Netspar

Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Marike Knoef

Essays on Labor Force Participation, Aging, Income and Health

PhD Thesis 2011-013

Essays on Labor Force Participation, Aging, Income and Health

Essays on Labor Force Participation, Aging, Income and Health

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op

vrijdag 2 september 2011 om 14.15 uur

door

Marike Geraldine Knoef

geboren op 15 januari 1983 te Deventer.

PROMOTORES: prof. dr. R.J.M. Alessie
prof. dr. A.H.O. van Soest
COPROMOTOR: dr. A.S. Kalwij

OVERIGE COMMISSIELEDEN: prof. dr. K.P. Goudswaard
prof. dr. M. Kalmijn
prof. dr. P. Kooreman
prof. dr. ir. J.C. van Ours

Preface

In my view, writing a Ph.D. thesis is comparable to running. Both are enjoyable and both are influenced by weather conditions. The weather may be calm, sometimes you have the wind in your back, and other times you need to run against the wind. In all circumstances advice and support help to reach the finish line. In this preface I would like to thank some special people that supported me to the finish line of this Ph.D. project.

During my Master thesis I had the pleasure to be supervised by Bertrand Melenberg (Tilburg University), Rob Euwals and Daniel van Vuuren at the Dutch Bureau for Economic Policy analysis (CPB). I still remember Daniel's advice which encouraged me to start a PhD project:"Make sure to keep your learning curve steep". Arthur van Soest, Peter Fontein and Marcel Das gave me the opportunity to start my Ph.D. project in combination with my work as a quantitative researcher at CentERdata for which I want to thank them a lot.

Foremost, I would like to express my sincere gratitude to my promotors Rob Alessie, Adriaan Kalwij and Arthur van Soest. During the process you all gave me very skilful advice and support. I have learned a lot from doing research with you and I deeply appreciate your effort and time. Also on a personal level working together was very enjoyable.

I would also like to thank the other members of the committee, Kees Goudswaard, Matthijs Kalmijn, Peter Kooreman, and Jan van Ours for taking the time to read the manuscript of this thesis. In particular, I wish to express my gratitude to Peter Kooreman, because writing a paper together was very inspiring and pleasant.

The Network for Studies on Pensions, Aging, and Retirement (Netspar) and Stichting Instituut GAK financed this Ph.D. research project. I thank both for their funding. In addition to their financial support, Netspar created a

stimulating research environment through their pension days and conferences, which greatly benefited this research. I also thank CentER for their financial support to present my work at conferences and workshops abroad.

Most of the data used in this thesis were provided by Statistics Netherlands. I thank the people of the 'Centrum voor Beleidsstatistiek' and Wim Bos for their advise on the Income Panel Study data.

I am grateful to my CentERdata colleagues for the pleasant atmosphere in which I worked on my thesis, for the 'vrijdagmiddagthee', and the yearly BBQ at Marcel's house. In particular I want to thank Hendri, my office mate, who was always willing to answer my LaTeX questions. The layout of this thesis would not have been the same without you. In addition, I sincerely thank my colleagues from the Department of Econometrics and Operations Research, from Netspar, and from the health-labor research group for discussions about research and other aspects of life. Furthermore, I thank my new colleagues at Leiden University, the first months with you have already been interesting. I am happy with the opportunity to extend the research in this Ph.D. thesis and to start a new project.

Nobody achieves goals without having fun along the way. I owe thanks to friends and family for all their interest. I am grateful to my paranymphs Meike and Irene, and I want to mention Anna, Barbera, Bastiaan, Chris, Frans, Geertje, Judith, Mark, Michiel, Marjon, Teun, Wendy, de Lochemse vriendinnen, Lilian, Lizette, Welmoed, other members from the former University choir 'Contrast', and the Saturday morning Tilburg Roadrunners.

I thank my parents, Jan and Gerrie, for all their love. Gerrie, the first ten years of my life I saw you studying day after day, year after year. Jan and I accompanied you to the train at 05.15am for your exams, and I could not imagine that I would ever study in my life as playing outside was much more fun. I could not have wished for a more stimulating example of eagerness and perseverance to learn. In addition to all the foregoing I need to mention Gijsbert. You bring a lot of happiness.

Marike Knoef
Tilburg, June 2011

Contents

Preface **vii**

1 Introduction **1**
1.1 Motivation 1
1.2 Research questions 3
1.3 Main findings 4

2 The Trend in Female Labor Force Participation: What can be Expected for the Future **9**
2.1 Introduction 9
2.2 Literature 12
2.2.1 The increase in female participation rates 12
2.2.2 The discouraged and added worker effect 15
2.3 Data 16
2.4 Empirical strategy 19
2.5 Estimation 21
2.6 Projected growth in two scenarios 30
2.7 Conclusion and discussion 34
2.A Extended estimation results 36

3 Pathways to Retirement and Cause-Specific Mortality Risks in the Netherlands **39**
3.1 Introduction 39
3.2 Data 43
3.2.1 Variable definitions and descriptive statistics 44
3.3 Cause-specific mortality model 48
3.3.1 Model estimation 50
3.3.2 Model identification 51

3.4 Estimation results 53
3.5 Summary and conclusions 60

4 The Effects of Cooperation; A Structural Model of Siblings' Caregiving Interactions 63
4.1 Introduction 63
4.2 Literature review 65
4.3 Structural model 69
4.3.1 Model specification 69
4.3.2 Modeling wage rates and remaining household income . 74
4.4 Data 76
4.4.1 SHARE 76
4.4.2 EU-SILC 81
4.5 Estimation results 83
4.5.1 Wage equation and remaining household income 83
4.5.2 Estimation results of the structural model 83
4.6 Two adult children 93
4.6.1 Cooperative and non-cooperative equilibria 94
4.6.2 Simulations 95
4.6.3 Interactions between siblings in SHARE 98
4.7 Conclusions 102
4.A Wage equations per country 104
4.B Logit equilibrium 108

5 Changes in the Income Distribution of the Dutch Elderly between 1989 and 2020: A Dynamic Microsimulation 109
5.1 Introduction 109
5.2 Developments in the Dutch income distribution 112
5.3 Data 113
5.3.1 Dutch income panel (IPO) 113
5.3.2 Population register (GBA) 122
5.4 Microsimulation model 123
5.4.1 Income equation 125
5.4.2 Differential mortality 128
5.4.3 Transitions in marital status and household composition 129
5.4.4 Transitions in labor market status 130
5.5 Estimation results 130
5.6 Simulation results 133
5.7 Conclusions 140
5.A IPO before and after the revision in the year 2000 142
5.B The construction of income 144
5.C Estimation method 145
5.D Estimation results of the transition models 146

5.D.1 Marital status . . . 146
5.D.2 Children . . . 147
5.D.3 Labor market status . . . 149
5.E Extended estimation results . . . 153
5.F Simulation results model specification 1 and 2 . . . 155
5.G Indexed growth in specification 1 and 2 . . . 157
5.H Theil decomposition . . . 158

6 The Association between Individual Income and Remaining Life Expectancy at the Age of 65 in the Netherlands 159
6.1 Introduction . . . 159
6.2 Data . . . 163
6.2.1 Variable definitions and descriptive statistics . . . 164
6.2.2 Differential mortality . . . 167
6.3 Mortality risk model . . . 169
6.3.1 Dynamic selection . . . 170
6.3.2 Estimation, identification, and empirical specification . . . 171
6.3.3 Monte Carlo simulations . . . 173
6.4 Empirical results . . . 174
6.4.1 Estimation results . . . 176
6.4.2 Simulation results . . . 178
6.5 Summary . . . 181
6.A Extended estimation results . . . 183
6.B Monte Carlo simulations . . . 190

Bibliography 191

1 Introduction

This thesis collects five studies that are related to population aging. Population aging is a predictable trend. However, unfortunately there are no standard recipes that describe how our society should deal with the social and economic consequences of aging. These five studies aim to contribute to the understanding of aging related issues. This introductory chapter first provides the motivation for this research (1.1). Second, I will present the research questions (1.2), after which section 1.3 summarizes the main findings of this thesis. All chapters in this thesis can be read independently.

1.1 Motivation

Population aging, caused by an increased life expectancy, the retirement of the baby boom generation, and a decreased birth rate, has a large impact on our society. Around 2040 the old age dependency ratio will have increased to around 55 (starting from 29 in 2010), meaning that in 2040 for every 100 persons of age 15–64 there will be about 55 persons over the age of 65.

Aging has raised concerns about the sustainability of the welfare state. Reforms are needed to keep the Dutch public and private pension system, and the health care system viable. In addition, the recent financial crisis has shown the vulnerability of the pension system and has increased the priority of pension reforms.

One of the directions to alleviate the pressure that aging has on our public finances, is to raise tax revenues by increasing the labor force participation of women and the 55+ population. In the last decades the female labor force participation has expanded considerably and the question arises what we can expect for the future. Chapter 2 of this thesis addresses this question. The labor force participation of the Dutch 55+ population declined during the 1980s and 1990s, due to generous early retirement schemes. Whereas the male participation rate was about 80% during the 1970s, it dropped to around 40% in 1995. As from the mid nineties the generosity of early retirement schemes has been diminishing. As a result, the labor force participation of the 55+ population increased again, up to about 64% of the 55–64 male population in 2010, and 51% of the total 55–64 population in 2010. The effectiveness and political support for new reforms aimed at increasing the labor force participation of the 55+ population, depends (among other things) on the health status of the elderly non-employed, which will be investigated in chapter 3.

A negative byproduct of an increase in the labor force participation of women and the 55+ population may be that the number of hours of informal care given by these groups decreases. This may be an unfortunate side effect as aging is expected to bring about increases in long term care spending[1] and labor shortages in the health care sector. The encouragement of informal care and labor force participation are two conflicting goals. For the development of effective policies, information is needed about the decision making process of adult children to provide informal care to their parents and to participate in the labor market. This decision is investigated in chapter 4. This chapter explicitly pays attention to the nature of the interactions between siblings.

The strong increase in the old age dependency ratio is problematic for our public and private pension system. Public pension benefits are paid on a pay-as-you-go (PAYG) basis. Without policy reforms, the public pension expenditures will increase from about 4.7% of GDP in 2009 to 8.8% in 2040. Occupational pension schemes, on the other hand, are based on capital funding and are therefore less sensitive to the expected increase in the old age dependency ratio. However, occupational pension schemes are more sensitive to investment risks, such as financial crises and inflation. The recent financial crisis had a large impact on the capital reserves of the Dutch occupational pension funds.

[1]The European Commission (2009) predicts that the long term care expenditures in the Netherlands increase from 3.5%-points of GDP in 2010 to 8.1%-points of GDP in 2060.

Also, the number of contributing members of the pension funds has declined, such that an increase in the pension premium has become a less effective measure to bear financial risks faced by the pension funds. The financial crisis, the relatively low number of contributing members, and the unexpected high rise in the life expectancy call for reforms to keep the Dutch pension schemes sustainable.

To assess the viability of proposed reforms to increase the sustainability of the public and private pension system, information is needed about the development of the income distribution of the elderly. Chapter 5 therefore describes the development in the income distribution for Dutch pensioners in the past, and predicts the income distribution of the Dutch elderly until 2020 in the absence of pension reforms.

When redesigning pension schemes, we have to be aware of the fact that low-income individuals have lower life expectancies than high-income individuals and therefore receive pension benefits for a shorter period of time. This has an adverse effect on the redistribution from the financially better to the financially worse off, which is the aim of public pension policies in many countries. For example, a rise in the statutory retirement age will reduce pension entitlements relatively more for low-income individuals than for high-income individuals. Chapter 6 examines differences in the remaining life expectancies of low- and high-income individuals after the statutory retirement age of 65 in the Netherlands.

Research questions 1.2

This thesis deals with five topics that were raised in the motivation of this thesis. This section presents the main research questions of this thesis more specifically.

The research questions addressed in chapter 2 are

2. Which factors were important in the increase of the female labor force participation in the Netherlands in the last two decades? And what can we expect for the future female labor force participation rates?

The research question handled in chapter 3 is
3. To what extent are pathways to retirement, such as early retirement and unemployment, associated with adverse health conditions?

Chapter 4 deals with the following research questions
4. How do adult children decide how much informal care they want to give to their parents, and how many hours they want to participate in the labor market? Do siblings make a cooperative or a non-cooperative decision to care for their parents? And what gains can be achieved from cooperation between siblings?

The research questions addressed in chapter 5 are
5. How did the income distribution of the elderly evolve over the last two decades? And how will the income distribution of the Dutch elderly develop in the coming decade in the absence of pension reforms?

The last research questions, that are answered in chapter 6, are
6. What is the association between income and the remaining life expectancy after the statutory retirement age of 65 in the Netherlands? And is it only individual income, or also the income of the spouse that is associated with mortality risk?

1.3 Main findings

This section provides the answers to the research questions presented in section 1.2.

Chapter 2 Female participation rates have increased considerably in recent years, from about 46% in 1992 to 59% in 2004 and 63% in 2010. Chapter 2 presents a decomposition of the female participation growth between 1992 and 2004, based on a binary age-period-cohort model. As age, period, and cohort effects cannot be disentangled, we have used several strategies to identify the model. The results show that most of the growth (40%) in the female participation rate between 1992 and 2004 can be attributed to the fact that female participation has become less sensitive to the presence of children (the negative effect of children has decreased). Secondly, about one quarter of the total growth is due to so called ‘unobserved cohort effects’, which reflect factors like social norms. They have increased the labor force participation of women up to the cohort born in 1955. The increased education level of women accounts for about one sixth of the total growth, and about 12% of the growth can be explained by the more favorable market conditions in 2004, relative to 1992. For the future, the growth of the female participation rate will slow down, in particularly because cohort effects have stabilized for generations born after 1955. Between 2005 and 2050 a further growth of 7–10%-points can be expected, which would alleviate the structural deficit caused by the aging of the population by about 1.0–1.5%-points of GDP.

Chapter 3 One of the most important concerns related to the labor force participation of the elderly, is that nonemployed individuals have health limitations that prevent them from remaining employed up to the statutory retirement age. Chapter 3 of this thesis investigates whether pathways to retirement such as early retirement and unemployment are associated with adverse health conditions, using cause specific mortality risks after the age of 65 as an objective measure for health status. We find that, compared to individuals who remain employed during the three years preceding the statutory retirement age, those who are early retired, unemployed, self-employed, or nonparticipating do not have a significantly different mortality risks for cancer, cardiovascular diseases, or other diseases. This finding suggests that the effectiveness of reforms of early retirement schemes and unemployment schemes to increase employment among older workers may, on average, not be adversely affected by health conditions. On the other hand, persons who receive disability benefits during the last three years before the statutory retirement age have significantly higher mortality rates after the age of 65. Their probability of dying before age 75, conditional on being alive at age 65, is almost two times higher than for persons who have been employed up to the statutory retirement age of 65. These results are derived from a competing risk model, that conditions not only on observed characteristics but also on unobserved individual specific effects. The model therefore takes into account that with increasing age, the sample becomes more selective in terms of both observed and unobserved characteristics. A methodological contribution of this chapter is the use of multiple causes of death statistics in the context of a competing risk model to allow the impact of unobserved individual characteristics to differ across the cause-specific mortality risks.

Chapter 4 Chapter 4 analyzes the decision making process of adult children to provide informal care to their parents. In the first part of the chapter a structural model is developed in which adult children without siblings maximize their utility over leisure, consumption and the amount of care their parents receive, subject to a time and a budget constraint. The model is estimated using two datasets from 12 European countries and estimates the preferences of adult children for consumption, leisure and informal care, without having to make assumptions about interactions between siblings.

The parameter estimates show that the preference for informal care increases when parents are in bad health. Women have higher preferences for informal care than men, and higher educated adult children have significantly lower preferences for informal care than lower educated adult children. Also cultural and institutional differences between countries play a role, and the (negative) wage elasticity of informal care supply appears to be small. The latter implies that fiscal policies that affect net wages have negligible effects on informal care (while they do influence labor supply). To increase informal care as well as labor supply a reduction of the geographical distance between adult children and their parents would be effective. For example, the social rent sector could weigh informal care in their assignment of houses, and senior houses could be built in residential areas.

In the presence of siblings, their choices also play a role in the caregiving decision. In the literature it has been emphasized that modeling family decisions as a bargaining process is important to improve our understanding of these decisions. The question arises whether this bargaining process between siblings is cooperative or non-cooperative. Do siblings maximize their total utility? Or do they maximize their own utility, given the behavior of their siblings? While in the literature it is often assumed that siblings behave non-cooperatively, chapter 4 presents a first attempt to identify the nature of the bargaining process between siblings. The results show that 71% of the siblings have a higher probability to behave non-cooperatively than cooperatively and 47% of the siblings even have a 10%-points higher probability to behave non-cooperatively than cooperatively. Examining the characteristics of cooperative and non-cooperative siblings reveals that two brothers have on average a 10.5%-points higher probability to behave non-cooperatively than two sisters. Pushing families into their cooperative equilibria increases informal care, but decreases labor supply.

Chapter 5 Also without pension reforms the income distribution of the elderly changes, as a result of changing household composition, developments in the labor market (e.g. the increased female labor force participation), productivity differences between cohorts resulting in income differences, differential mortality, and increased longevity. Chapter 5 describes the income distribution of the elderly in the past, and predicts the income distribution of the elderly until 2020 in the absence of pension reforms, using a dynamic microsimulation model. The results show that equivalized household income of the elderly in the age group 65–90 is expected to increase on average by 0.5% per year for the 10th percentile, 1.2% for the median, and 1.0% for the 90th percentile. Income inequality grows in the lower part of the income distribution, but declines in the upper part of the distribution. The contradictory movements in the lower and upper part of the distribution underline the importance of investigating the whole income distribution, here achieved using a microsimulation model, instead of only analyzing an inequality index such as the Gini coefficient. Our microsimulation model deviates from traditional microsimulation models by explicitly paying attention to the modeling of income shocks and the persistence of these shocks, and taking into account that younger cohorts are different than older cohorts, for example with regard to female employment and divorce rates.

Chapter 6 Low-income individuals live shorter than high-income individuals. This implies that low-income individuals have a lower rate of return from uniformly priced pension plans than high-income individuals. Chapter 6 quantifies the association between income and the remaining life expectancy after the statutory retirement age of 65 in the Netherlands. The results are obtained using a mortality risk model that allows for dynamic selection based on both observed and unobserved characteristics. This means that we take into account that the population at risk changes with age, and that only those with relatively advantageous observed and unobserved characteristics survive (survivorship bias). The results show that individual income is about equally strong and negatively associated with mortality risk for men and women. Furthermore, only for women, spouse's income is weakly associated with their mortality risk.

For both men and women, the difference in the remaining life expectancy at age 65 between low-income individuals with only a public retirement pension and high-income individuals with twice the median income is about 2.5 years. This difference adversely affects the redistribution of income from high-income individuals to low-income individuals and underscores the importance to allow for a retirement window, that is part of the proposed pension reforms in the Netherlands. A retirement window[2] increases people's free choice and can mitigate the adverse income redistribution effects that result from differential mortality.

[2]E.g. a retirement window in which people can choose their retirement age between 65 and 67, and in which public pension benefits are adjusted in an actuarially fair way.

2 The Trend in Female Labor Force Participation: What can be Expected for the Future?

This chapter is published as a paper in Empirical Economics (Euwals, Knoef, and Van Vuuren 2011).

2.1 Introduction

Over the last decades, many countries experienced an increase in the labor force participation of women. The Netherlands are an exceptional example, as female participation more than doubled, from 31% in 1975 to 69% in 2006 (OECD, various years). This development reflects a major change in the Dutch society and has several dimensions. First, an important social dimension is that women are becoming more integrated into formal production. Although women lag behind men considerably in terms of (full-time) employment and wages, like in many other countries, at least they are catching up. Second, an important economic dimension is that an increase in the female contribution to formal production leads to higher economic growth. The high growth rates of the Dutch economy at the end of the 1990s can be attributed partly to the substantial increase in female participation. Third, the developments have a fiscal and demographic dimension. It is widely believed that an increase in the participation rate contributes to the fiscal sustainability of the welfare state, which is under pressure due to the aging of the society. An important portion of the foreseen fiscal gap could be prevented by an increment in the labor force participation (Aaberge et al., 2004, Apps, 1991, Van Ewijk et al., 2006).

The first goal in this chapter is to decompose the growth in the female participation rate over the period 1992–2004. Do younger cohorts have a larger probability to be employed than older cohorts, holding age and year effects constant? Did the favorable economic conditions in the late 1990s encourage women to participate in the labor market? Our modeling approach allows us to make a full decomposition of the observed growth during the period mentioned. Our second aim is to determine which factors will remain important in the coming decades. While some factors may not be as important for future growth as they have been in the past, others are likely to remain important. The effects of the various factors on the future development of the participation rate are quantified in two scenarios. As an aside, we will also give some indication on the effect of the prospected participation increase on the fiscal sustainability of government finances in the Netherlands. Typically, most studies on the impact of aging on government finances do not assess the partial impact of participation growth on fiscal sustainability, but rather take as given one central projection (see e.g. Cournède and Gonand, 2006, Roeger, 2002, Rother et al., 2002).

To answer these research questions, we estimate a binary choice model for the labor force participation of women born between 1925 and 1986 on the basis of the Dutch Labor Force Survey 1992–2004. We employ two different identification strategies to disentangle the participation growth into age, period, and cohort effects. According to our interpretation, the age effect includes individual life-cycle decisions, like the timing of education and marriage. Period effects include cyclical and instantaneous effects, e.g. effects of policy changes. Cohort effects, that is different participation rates among different generations of women, are linked to societal changes in the orientation towards paid employment. These cohort effects can for instance be related to a higher educational attainment or lower fertility of younger generations.[1] Apart from such 'observed' cohort (and period and age) effects, the model also allows for 'unobserved' cohort effects which are not directly related to variables included in the model. These unobserved cohort effects are mainly related to the evolvement of social norms or the availability of oral contraceptives, or a combination of both. Social norms are an appealing explanation as sociological research on social norms and attitudes with respect to the combination of employment

[1]Note that there is some ambiguity in the distinction between period and cohort effects. Cohort effects can sometimes be indirectly linked to earlier policy changes.

and family care responsibilities finds an almost identical development over the cohorts (SCP/CBS, 2006). Furthermore, birth control may also have played a role as oral contraceptives became available in the years the cohorts born in the 1950s became mature (Goldin and Katz, 2002).

Estimation results show that both the observed and unobserved cohort effects have been crucial in spurring the female participation rate. Time effects have also played an important role during the 1990s. The negative relation between the unemployment rate and the participation rate suggests an 'encouraged worker effect', i.e. the favorable market conditions during the 1990s have induced many females to participate in the labor market. An important finding is that observed factors (such as education and household situation) can fully 'explain' participation growth of generations born after 1955, but that unobserved cohort effects play an important role for the older cohorts. This means that unobserved factors, like social norms, have only been important for cohorts born before 1955. This finding is consistent with both sociological studies, which find important shifts in societal preferences until the 1950s cohorts but not thereafter, and with studies focusing on oral contraceptives as an explanation for increasing participation rates of women. The outcomes give us some important indications about the future prospects of the labor force participation of women. To make this point explicit, we construct long term scenarios. Apart from a 'basic scenario', where attitudes towards the combination of paid work and care for children stabilize, we construct an 'emancipatory scenario', in which these attitudes shift in favor of combining paid work with raising children. It is not made explicit though whether this shift is the result of a policy change or due to a further change in social norms. The scenarios predict that female participation rates grow with a further 7 and 10%-points, respectively. Using these scenarios, we estimate that the rising female participation rates compensate for about one third of the total fiscal sustainability gap[2] in the Netherlands.

The remainder of this chapter is organized as follows. Section 2.2 discusses some relevant literature on female labor supply. Section 2.3 discusses the data. Section 2.4 discusses the empirical strategy, and in particular the identifica-

[2]The fiscal sustainability gap is defined as the structural deficit given current institutions and future demographics. As a result of the aging population, government expenditures will rise much faster than revenues.

tion strategies used. Section 2.5 discusses the empirical findings, section 2.6 discusses future projections and the consequences for fiscal sustainability, and section 2.7 concludes.

2.2 Literature

In many countries female labor force participation has increased substantially over the last few decades. The literature has paid a lot of attention to this development, resulting in a large number of empirical studies. The approach in these studies varies from structural modeling of financial incentives and life cycle decision making (for an overview, see, e.g. Blundell and MaCurdy, 1999) to the historical analysis of changing life courses of women (see, for example, Goldin, 2004, 2006). Our study will take a reduced form approach that is related to the last type of studies. This section will therefore not discuss the extensive structural literature, but instead focus on studies which describe changes in labor market behavior over a longer time period on the basis of panel data and repeated cross sections. labor market behavior due to cyclical changes – the discouraged worker and added worker effect – plays an important role in our model, and is therefore briefly addressed at the end of this section.

2.2.1 The increase in female participation rates

Female labor market participation increased substantially in many countries over the past decades (table 2.2.1). The Netherlands stands out with an increase of about 40%-points over de last three decades, together with Spain with an increase of almost 30%-points over the same period. The empirical literature contains many studies on the increase in particular countries, and we discuss studies for the Netherlands, Germany, the UK and the US. We do not discuss the studies on Scandinavian countries, as these mostly focus on the high full-time employment rate of women (see, for example, Pfau-Effinger, 1993, Sundström, 1991).

The labor market participation rate of Dutch women started to increase from the 1970s onwards. Hartog and Theeuwes (1985) used the Terms of Employment Survey 1979 to investigate the labor supply behavior of women.

Table 2.2.1: Labor force participation (%), women 15–64 (source: OECD Statistical Database)

	1977	1987	1997	2007
Denmark	65	77	75	76
France	53	56	60	64
Germany	51	54	62	70
Italy	38	43	44	51
Netherlands	32	49	62	73
Spain	33	38	48	62
Sweden	70	79	75	78
United Kingdom	56	63	67	68
United States	56	66	70	70

They conclude that wage growth contributed substantially to the explanation of the increase in participation in the years after the Second World War. Note that at that time the female participation rate was clearly below the average of the OECD countries.

From 1979 until 1987 the female participation rate has increased to about 50%, and Groot and Pott-Buter (1993) use the Supplementary Benefits Surveys 1979 and 1987 to investigate this increase. They conclude that changes in preferences must have induced the increase in female participation. Note that in these years real wage growth was low due to the economic crisis and the Dutch wage moderation policy.

Female participation kept increasing during the 1990s, and Cörvers and Golsteyn (2003) use the Socio-Economic Panel 1994–1999 to investigate this increase. Like the previous authors, they find that preferences must have played an important role. During that period wage growth was still low due to wage moderation. Henkens et al. (2002) use the Housing Demand Surveys 1989/90 and 1998/99 to compare married and cohabiting women. They find that in particular the participation of married women has increased. This again hints at a change in preferences, and in particular the preferences of married women.

The female participation rate in Germany increased by almost 20%-points during the last three decades. Fitzenberger et al. (2004) and Fitzenberger and Wunderlich (2004) use the Micro-Census to investigate the increase. The studies find an increasing time trend for low and medium skilled women, implying an absence of an increase over cohorts. Only for high-skilled women they do find a cohort effect. Women entering the labor market after 1975 have a higher probability to participate than older cohorts. The second study furthermore concludes that the increase in the participation rate is concentrated in part-time employment and, most importantly, this increase is mainly due to the changing age and skills composition.

Both the UK and the US experienced an increase in female labor market participation of about 10%-points during the last three decades. Excellent census data made early studies possible. In both countries, economists were particularly interested in the question whether the increase in real wages could explain the increase in female participation. For the UK, Layard et al. (1980) used cross-section data from the UK Household Survey 1974 in order to estimate the wage elasticity, and concluded that wages explained about a third of the increase in female participation during the period 1973–1977. Joshi et al. (1985) turned to repeated census and survey data for the period 1850–1980.[3] They isolate a clear cohort effect. They conclude that wages only explain a minor part of the increase in female participation. The authors offer some tentative explanations, such as falling prices of domestic services, changing fertility, and long-term changes in the roles women see for themselves in life.

On the basis of US census data 1890–1980, Smith and Ward (1985) conclude that rising real wages account for 60 percent of the total growth in the female labor force after 1950 in the US. Nevertheless, they are not able to explain why participation did not increase earlier in history despite the fact that wages did increase too. They argue that other factors must have been important as well, and tentatively mention fertility, schooling, and changing attitudes towards women's work. Coleman and Pencavel (1993) use census data and the Current Population Survey (CPS) for the period 1940–1988, and they report strong cohort effects in female participation. Female working hours did not change substantially over the different cohorts, and the authors conclude that gender differences in work behavior are becoming less manifest in the US. Pencavel (1998) uses the CPS 1975–1994 and again finds that in particular participation has increased substantially, while working hours have increased only slightly. He furthermore finds that in particular participation

[3] The special issue of the *Journal of Labor Economics* contains several articles that use repeated cross sections or repeated census data, including Smith and Ward (1985) for the US, Hartog and Theeuwes (1985) for the Netherlands, and Colombino and De Stavola (1985) on Italy.

of high-skilled women has increased. Attanasio et al. (2008) use the PSID 1969–1998 to calibrate a life-cycle model for three generations of women. They conclude that in particular shifts in the costs of children relative to life time earnings are the most likely explanation for the strong increase in labor supply over the generations of women.

Another strand of the US literature discusses changes in life cycle labor supply behavior of women over generations. Goldin (2004), Goldin (2006), Fernandez et al. (2004) and Fogli and Veldkamp (2007) discuss theories in which younger generations of women learn from older generations. In particular the last study explicitly models and tests the slow process in which each generation updates their parents' beliefs on maternal employment by observing the children of employed women. The results offer an explanation for the slow S-shaped rise in maternal employment in the US. Goldin and Katz (2002) claim that birth control technology ('the pill') played an important role as well. On the basis of differences in the date of first availability between states in the US, they find an impact of the pill on educational attainment and employment over successive generations of women.

Overall, conclusions on the Netherlands seem to direct towards changes in labor supply preferences of women. A similar result is found for highly educated women in Germany, and also for the UK and the US there is evidence in this direction. The tendency in preferences of Dutch women is addressed by Vendrik (2001), who employs the theory of social custom of Akerlof (1980) to describe the potential importance of evolving norms within society for the increase in labor market participation of women. This explanation fits in the US literature. On the basis of this finding from the literature, we may expect to find an 'unobserved cohort effect' in our empirical analysis. The size and duration of this effect may play an important role in the long term development of female labor force participation.

The discouraged and added worker effect 2.2.2

The business cycle affects many participants and potential participants to the labor market at the same time. For modeling purposes (see section 2.4) it is important to control for the impact of the business cycle on labor supply behavior. The business cycle effect may take place through the 'discouraged worker effect' and/or through the 'added worker effect'.

The discouraged worker hypothesis states that potential participants withdraw from the labor market because they believe that their chances to find suitable employment are low because of an unfavorable labor market situation. Several studies using time series analysis (Benati, 2001, Darby et al., 2001, Gregg, 1994, Tachibanaki and Sakurai, 1991) indeed show that participation displays a pro-cyclical pattern.

The added or additional worker hypothesis states that unemployment of the husband induces the wife to participate. As unemployment is high during economic downturns, it counteracts the discouraged worker effect. The existence of this effect is however questionable, and may at best be small. No evidence for an added worker effect is found on the basis of the PSID 1976 and 1982 (Maloney, 1991, 1997). On the basis of displaced workers in the PSID 1968–1992, Stephens (2002) finds that the long-run increase in the wife's labor supply accounts for about 25% of the husband's lost income. On the basis of the European Panel Household Survey 1994–1996, Prieto-Rodriguez and Podriguez-Gutierrez (2000) find that only in a few countries the participation of wives is related to the husbands' labor market status.

2.3 Data

The data used in this chapter are from the Dutch Labor Force Survey 1992–2004 (DLFS; in Dutch 'Enquête Beroepsbevolking'). The DLFS is a survey conducted among persons living in the Netherlands, with the exception of those living in institutions like nursing homes and prisons. Every year, a random sample of about 1% of the Dutch population aged 15 and older is interviewed. A new random sample is drawn every year, so that we do not observe multiple observations for a given individual. The DLFS collects information on the individual labor market situation and on individual and household characteristics.

From 1992 onwards, definitions and methods in the DLFS have changed radically, which means that prior years could not be used in our analysis. In the new definition, only persons working at least 12 hours a week and persons actively searching for a job of at least 12 hours a week belong to the labor force.

Table 2.3.1: Summary statistics (source: DLFS, 1992–2004)

	Mean	Min.	Max.	# Obs.
Participate	0.53	0	1	516298
Position in household				
Married	0.60	0	1	516298
Cohabiting, been married	0.02	0	1	516298
Cohabiting, never been married	0.11	0	1	516298
Single, been married	0.10	0	1	516298
Single, never been married	0.09	0	1	516298
Living with parents	0.07	0	1	516298
Other	0.01	0	1	516298
Children				
No children	0.57	0	1	516298
Age youngest child 0–3	0.12	0	1	516298
Age youngest child 4–11	0.13	0	1	516298
Age youngest child 12–17	0.09	0	1	516298
Two children < 18	0.16	0	1	516298
Three or more children < 18	0.07	0	1	516298
Both children < 18 and ≥ 18	0.05	0	1	516298
Only children ≥ 18	0.10	0	1	516298
Education woman[a]				
Primary	0.14	0	1	516298
Lower secondary	0.27	0	1	516298
Higher secondary	0.39	0	1	516298
Tertairy	0.20	0	1	516298
Education male partner[a]				
Primary	0.08	0	1	380686
Lower secondary	0.15	0	1	380686
Higher secondary	0.30	0	1	380686
Tertairy	0.18	0	1	380686
Age	41	18	67	516298
Period (year)	1998	1992	2004	516298
Cohort (year of birth)	1957	1925	1986	516298

[a] The education levels are defined as follows: 'Primary' = no secondary education completed (just primary shool); 'Lower secondary' = lower vocational or general school completed (in Dutch: VMBO); 'Upper secondary' = advanced vocational or general school completed (in Dutch: MBO, HAVO, VWO); 'Tertiary' = academic or vocational colleges completed.

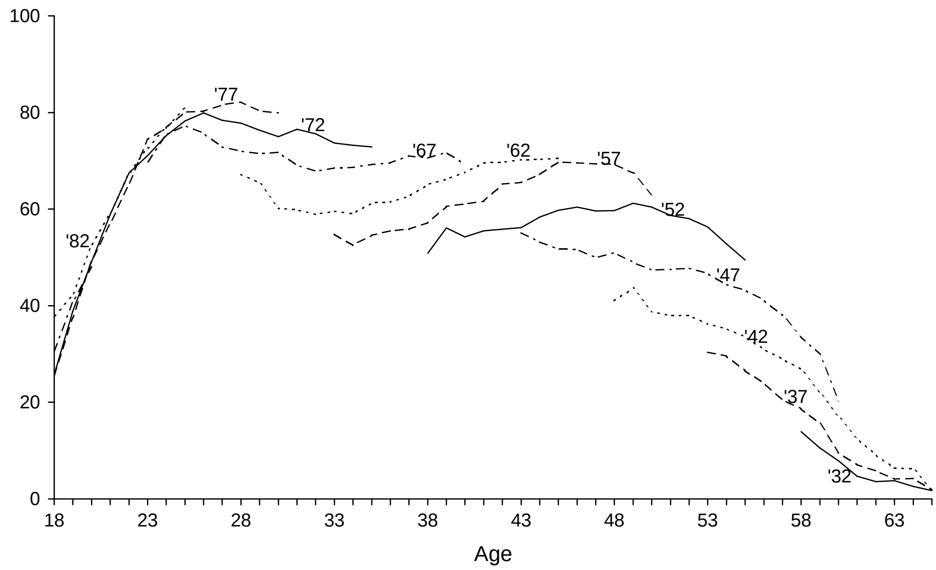
We use this definition of the gross participation rate, as it is the official definition used in the Netherlands. In 2004, about 6% of the Dutch women between the ages of 15 and 64 were working positive hours, but less than 12 hours per week (Euwals et al., 2006). That is, the participation rate would be about 6%-points higher according to the internationally accepted definition of working at least 1 hour per week. The average participation rate in our sample equals 53%.

The resulting data set contains 516,298 observations of women, which amounts to nearly 40,000 observations per year. The oldest cohort was born in 1925, and the youngest cohort was born in 1986 (table 2.3.1). Each cohort (age group) contains about 8,000 (10,000) individuals. The number of observations is somewhat lower for the oldest age categories and cohorts due to mortality. Furthermore, note that the smallest category ('Other household member') still counts more than 5,000 observations. More than half the sample has no children living at home. About one out of eight women have at least one child younger than 4 years old living at home. Note that these and other variables on children in the household are not mutually exclusive.

We add education-specific delayed unemployment rates to the individual records, so that we will be able to assess the combined impact of discouraged and added worker effects (see the previous section). The labor market situation given educational attainment is thus represented by the one-year delayed unemployment rate given that education level. We have chosen the one-year delayed unemployment rate because non-participating women are not likely to have up-to-date information on their labor market prospects.

The observed participation rates by cohort and age are shown in figure 2.1. Participation rates are increasing steeply until age 25. Until that age the youngest three cohorts show no difference in participation rates. As from age 25 a gradually decreasing pattern appears. From the overlapping segments at a given age it can be seen that increments of about 10%-points between subsequent five-year cohorts are not uncommon. These 'jumps' are combined cohort and period effects. Between age 25 and 35 there appears to be a 'motherhood dip'. This dip seems to become less strong for younger cohorts. In the following section we investigate whether this is related to a drop in the fertility rate or a changing attitude towards the combination of working and caring for children.

Figure 2.1: Female participation rates by birth cohort and age (%)



Note: Cohorts in 5-year groups, from cohort born in 1980–1984 (left in figure) to cohort born in 1930–1934 (right in figure). Source: DLFS, 1992–2004.

Empirical strategy 2.4

In this section, we specify the statistical model that is used to estimate the determinants of female labor force participation. Indicating individual i and (discrete) time t by corresponding subscripts, our model specifies the propensity to participate in the labor market as

$$p^*_{it} = \beta_0 + \beta_1 x_{it} + g_a(a_{it} \mid \theta_a) + g_c(c_i \mid \theta_c) + g_t(t \mid \theta_t) + \varepsilon_{it}, \qquad (2.1)$$

where x_{it} is a vector of control variables, a_{it}, c_i, and t are age, cohort, and year effects, respectively, and corresponding transformation functions are denoted by g. We specify the probability of observing individual i participating in the labor market at time t as a binary Logit (i.e. we assume that ε_{it} follows a standard logistic distribution)

$$\Pr(p_{it} = 1 \mid x_{it}, a_{it}, c_i, t) = \Pr(p^*_{it} > 0 \mid x_{it}, a_{it}, c_i, t) = \frac{\exp\left(\beta_0 + \beta_1 x_{it} + g_a(a_{it} \mid \theta_a) + g_c(c_i \mid \theta_c) + g_t(t \mid \theta_t)\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{it} + g_a(a_{it} \mid \theta_a) + g_c(c_i \mid \theta_c) + g_t(t \mid \theta_t)\right)}, \qquad (2.2)$$

where p_{it} equals unity if individual i participates at time t, and zero otherwise.

A well-known complication in (2.1) is that not all parameters can be identified whenever all transformation functions g_a, g_c, and g_t contain a (parameterized) linear term. The reason is that the following identity holds for any individual i at time t

$$c_i + a_{it} = t. \tag{2.3}$$

Whenever two terms in (2.3) are known then the third is known as well. A large literature going back to the 1970s has examined the problem of identifying age, period, and cohort effects. For instance, Hall (1971) identified his model by assuming that the two most recent cohorts were identical. Essentially, assumptions are needed to identify the model. To avoid arbitrary results, these assumptions have to be based on some a priori knowledge, for instance from economic theory. In the present case, we expect that period effects largely correspond with discouraged and added worker effects (see section 2.2). It is therefore conceivable that period effects can be well represented by a variable which is directly related to both these theories: the unemployment rate. In the literature this kind of procedure is often called the proxy variable approach, and is e.g. applied in Smith and Ward (1985) and Kapteyn et al. (2005). In practice, this means that we specify $g_t(t \mid \theta_t) = \theta_t UNEMP_t$, where $UNEMP_t$ is the unemployment rate per education group at time t.

A second approach is based on the a priori expectation that increases in participation rates over subsequent cohorts are diminishing. This is not so much a theoretical identification assumption but rather a technical one, as participation rates are likely to grow less fast when the strict upper bound equalling unity is approached. In this respect, we assume cohort effects to follow a logarithmic pattern over time. This identification approach, which is often called the 'functional form approach', assumes that the g-functions follow a prespecified functional form for which the model parameters in (2.1) are identified. This kind of strategy was e.g. used in Fitzenberger et al. (2004). In practice, this means that we specify $g_c(c_i \mid \theta_c) = \theta_c \ln(c_i)$.

A final approach is to define the g-functions as step functions of the form $g(y) = \sum_{k=1}^{K} \alpha_k \mathbf{1}\{y = y_k\}$, where K is the total number of categories for variable y, and $\mathbf{1}\{A\}$ is the indicator function for event A. Additional restrictions on the parameters (α_k) are imposed in order to achieve identification. This widely used approach, which was first introduced by Mason et al. (1973), was however criticized for the arbitrariness of the imposed restrictions, and the fact that it is not testable whether the restrictions are valid. We will therefore not apply this approach in the current chapter.[4]

Having estimated a model specification as described above, we may decompose the growth in female participation for different years. Denoting the probability in (2.2) by q_t (and omitting the individual subscript), we compute the marginal effect of variable x_j as

$$\frac{\partial q_t}{\partial x_j} = q_t(1-q_t)\beta_j,$$

and approximate the change in participation as a result of a change in factor x_j at time t by

$$e_{jt} = \overline{q}_t(1-\overline{q}_t)\beta_j \Delta \overline{x}_{jt}, \tag{2.4}$$

where $\overline{q}_t$ denotes the predicted probability that an average female is participating at time t, and $\overline{x}_{jt}$ denotes the average value of covariate j at time t. We will employ this formula in the next section when making a decomposition of the aggregate growth in female labor force participation during the period 1992–2004.

Estimation 2.5

This section discusses the estimation results of the specifications explained in the previous section. The coefficients reported in table 2.5.1 and the appendix represent the impact of a variable on the probability of participation. For example, the positive coefficients for cohabiting women imply that this category is more likely to participate in the labor market than married women. We start with an interpretation of the estimation results for Model I, and then proceed with the results for the alternative specification (Model II).

The first specification uses the proxy variable approach, where period effects are proxied by the variable 'delayed unemployment rate' (Model I). The delayed

[4]We have experimented with this methodology and indeed found the results to be sensitive for the identification restrictions.

Table 2.5.1: Probability of participation in the labor market, estimation results for model I (proxy variable approach) and model II (functional form approach)[a]

	Model I		Model II	
	Coef.	S.e.	Coef.	S.e.
Position in household				
Married				
Cohabiting, been married	0.50	0.03	0.50	0.03
Cohabiting, never been married	0.91	0.03	0.91	0.02
Single, been married	0.25	0.05	0.25	0.04
Single, never been married	0.07	0.06	0.07	0.04
Living with parents	0.51	0.08	0.52	0.05
Other	0.20	0.07	0.20	0.06
Children				
No children				
Age youngest child 0–3	-1.66	0.11	-1.66	0.03
Age youngest child 4–11	-1.27	0.11	-1.26	0.03
Age youngest child 12–17	-0.81	0.11	-0.79	0.03
Only children ≥ 18	-0.34	0.04	-0.34	0.01
Two children < 18	-0.42	0.02	-0.41	0.01
Three or more children < 18	-0.89	0.02	-0.89	0.01
Both children < 18 and ≥ 18	-0.11	0.04	-0.10	0.02
Interaction: single & child < 18	-0.14	0.03	-0.14	0.02
Interaction: single & child ≥ 18	0.32	0.04	0.32	0.03
Interaction: lower secondary & child < 18	-0.05	0.04	-0.04	0.02
Interaction: upper secondary & child < 18	-0.02	0.06	0.00	0.02
Interaction: tertiary & child < 18	0.37	0.05	0.39	0.03
Interaction: period & child < 18	yes	**	yes	**
Education				
Education woman	yes	**	yes	**
Education male partner	yes	**	yes	**
Interaction: woman & male partner	yes	**	yes	**

[a] Estimates for the education and interaction effects, and the unobserved year and age effects in Model II can be found in the appendix. The unobserved cohort and age effects in Model I are shown in figures 2.2 and 2.3. The '**' indicate that corresponding dummy variables are jointly significant at the 5% confidence level (F-test).

[b] Note that in computing standard errors, we have taken into account that the unemployment variable is not observed at the individual level, but rather per year and per level of education. The standard error for the variable would otherwise be underestimated.

Table 2.5.1: Probability of participation in the labor market, estimation results for model I (proxy variable approach) and model II (functional form approach) (*continued)*

	Model I		Model II	
	Coef.	S.e.	Coef.	S.e.
Age				
Dummies	yes	**	yes	**
Period (year)				
Dummy variables			yes	**
Unemployment	-0.05	0.01		
Cohort (year of birth)				
Dummy variables	yes	**		
Ln(cohort-1924)			1.27	0.06
Log likelihood	-272180		-272607	
Pseudo-R^2	0.239		0.237	
# Observations	516298		516298	

unemployment rate is known by education level (see section 2.3). The estimation results show that women are less likely to participate when the level of unemployment is high, implying that the discouraged worker effect is stronger than the added worker effect. This is in line with the empirical literature discussed in section 2.2. Furthermore, being married is negatively related to the labor force participation of women. On the other hand, women who are part of an unmarried couple have a relatively high probability to participate.

Children have a negative effect on female participation, and this negative impact is particularly strong in the presence of children under 4 years of age. Participation goes up once the youngest child attends primary school, and again when it attends secondary school (in the Netherlands this usually occurs at the ages of 4 and 12, respectively), and once again when it reaches the age of 18. The negative participation effect of having children increases in magnitude when more children are present.

From the interaction terms between having children and the year of observation (shown in the appendix), we learn that the negative effect of raising children has decreased importantly between 1992 and 2004. Participation rates of women with children are rapidly converging to that of women without children. By 2004, the negative participation effect of a youngest child between the ages of 12 and 17 (-0.81) is canceled out by this trend effect (0.78). The decrease of the negative child effect may be the result of the increase in the availability and affordability of child care facilities, which is partly the result of higher subsidies from the government. Another explanation may be a change in attitudes towards the combination of paid employment and caring for children.[5] Note that the reported coefficients should not be interpreted as causal effects, as both participation and children (as far as it is a choice) may be the result of a simultaneous decision. Therefore, the causal effect of children on participation may be smaller in magnitude than the point estimates reported.[6]

A high level of education leads to more participation. Within the context of our model, the individual's educational attainment principally serves as a proxy variable for her wage rate. The control for educational attainment means that we do take into account the increase in real wages due to an increase in the level of education. We are however not able to correct for real wage increases given the level of education, as we have no data available on this. To have some idea of the concerning effect we collected some aggregate data on real wage trends (Loon Structuuronderzoek; LSO). We find that between 1997 and 2002 real gross wages have on average increased by 0.5% for elementary education, and 8.3% for university education.[7] In our model specification it is most likely that the increased participation level due to wage increases per education level is part of the cohort effect, because year effects are correlated with unemployment and not with structural wage changes.

The level of education of the partner also plays a role in the decision to participate in the labor market. For primary to upper secondary education (of the partner) a higher level of education of the partner has a positive effect on females' participation rates.

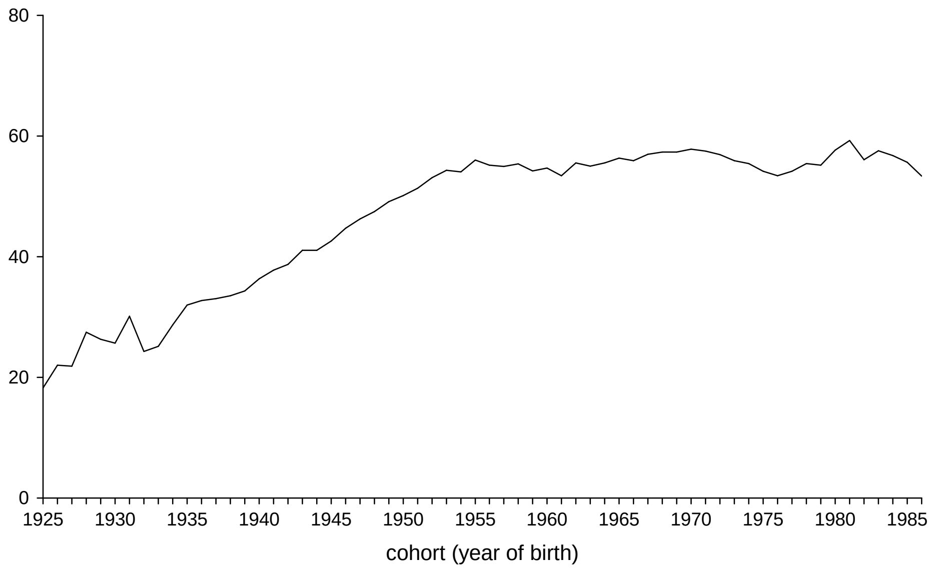
The estimated unobserved cohort and age effects are shown in figures 2.2 and 2.3. These figures contain the participation probabilities in equation (2.2) with explanatory variables at their sample average values, except for the cohort and age dummy variables. In this way the unobserved cohort and age effects can be interpreted in terms of participation growth. Note that each birth year / age

[5]Note that compared to other countries, Dutch women traditionally have a rather negative attitude towards the use of child care facilities (Kremer, 2005).

[6]In addition, estimates of the effect of children on the participation decision typically suffer from omitted variable bias, as having children is likely correlated with unobserved taste factors (Nakamura and Nakamura, 1994).

[7]Separate figures for men and women are not available for this time period.

Figure 2.2: Estimated profile of unobserved cohort effects, Model I (%)

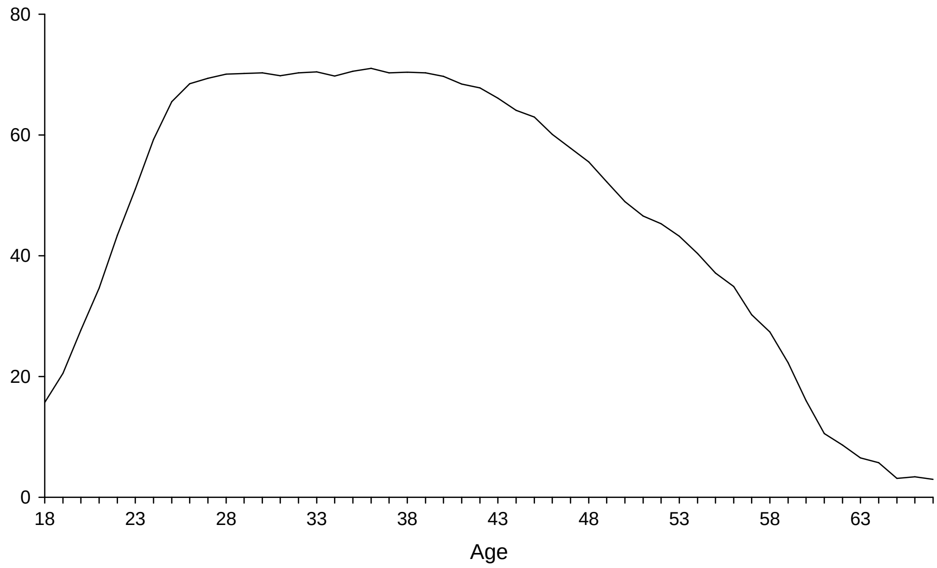


Note: Estimated probabilities to participate, where the mean of the data is taken for all characteristics except for the cohort dummy variables.

is represented by a unique dummy variable, implying that both cohort and age effects are estimated with a maximum degree of flexibility. From the first figure it can be seen that the unobserved cohort effect has increased almost linearly over the cohorts born between 1935 and 1955. This means that the variables included in our model cannot ‘explain’ the entire participation growth for these cohorts, and hence, other forces are at work. On the other hand, the unobserved cohort effect remains remarkably stable for cohorts born after 1955, implying that the included variables on educational attainment, children, and household situation (and interactions) are able to ‘explain’ increasing participation rates for younger cohorts. This fits well into the story of evolving social norms and better birth control opportunities (see section 2.2). The pattern we find is strikingly similar to results from sociological research on social norms. On the basis of repeated surveys from 1970 to 2004, SCP/CBS (2006) find that the generations born before the 1950s found it less and less problematic that women with children are employed, while the more recent generations of the 1970s and 1980s nowadays have about the same opinion as the generation born in the late 1950s.

From figure 2.3, it can be seen that unobserved age effects increase until the age of 25. They are more or less constant between ages 25 and 40, and decrease from the age of 40 years. Comparing this figure with figure 2.1 we see that the child-related variables make the dip disappear, so that the 'motherhood dip' around the age of 30 is indeed explained by the presence of children. It is however remarkable that the age profile already starts decreasing around the age of 40. Complementarity of leisure time with that of an older partner could be an explanation for relatively early retirement of women (Hurd, 1990, Schirle, 2008). It may also be the case that some delayed effect is taking place. Women who have lost attachment to the labor market while taking care of young children may find it difficult to reignite their careers, and in the end decide to withdraw from the labor market altogether. As our model specification does not include a delayed effect of having children, this effect is part of the unobserved age effect.

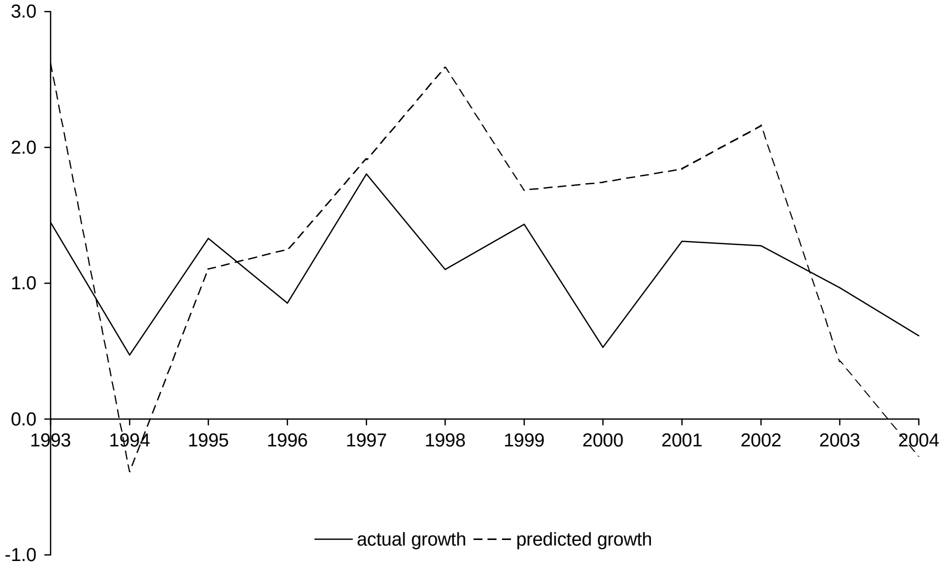
Figure 2.3: Estimated profile of unobserved age effects, Model I (in %)



Note: Estimated probabilities to participate, where the mean of the data is taken for all characteristics except for the age dummy variables.

In figure 2.4, we compare the actual growth of female participation rates with model predictions. For most years the predicted growth rates are quite accurate; for 1993, 1998, and 2000 the model overestimates growth by more than one %-point, and for 1994 and 2004 the model underestimates growth by nearly a %-point. During the years 1998–2002 the model underestimates participation growth, which could well be related to the strong economic upturn during that period.

Figure 2.4: Actual and predicted growth in the female participation rate (in %-point)



Note: predicted growth rates are based on marginal effects, i.e. first order approximations. Deviations between predicted and actual growth rates may thus be the consequence partly of second and higher order effects.

Table 2.5.1 also reports the results for the second specification, which is based on the functional form approach (Model II). This specification postulates that the unobserved cohort effects follow a logarithmic pattern over time. An attractive feature of this functional form is that its derivative tends to zero as time passes by, implying that the increase in participation rates over subsequent cohorts becomes smaller and smaller, which is expected. Moreover, results from the previous specification suggest that this functional form is not a strong assumption. As can be read from the table, parameter estimates for both models are mostly qualitatively the same. Results on the unobserved age and cohort effects are similar as well, except for the cohort effect being obviously much more smooth compared to the effect in figure 2.2 as a result of the functional form assumption. The pseudo-R^2 of Model II equals 0.2374.[8] Compared to the score of 0.2386 in Model I this is a minor reduction, given that Model I contains 60 more parameters than Model II.

The decomposition of participation growth for the years 1992–2004 based on Model I is reported in table 2.5.2. The respective elements of the decom-

[8]The pseudo-R^2 is defined as 1 - [the log likelihood of the model] / [the log likelihood of an alternative model which contains only an intercept].

position have been computed as in equation (2.4). That is, average values for all variables are determined both in 1992 and 2004, and the difference is used to predict the change in the female participation rate. To interpret results correctly, we take unemployment as an example: 1.6% is the change in participation for an 'average woman' in 1992 which sees the unemployment rate decrease to the 2004 level. As the initial decomposition does not add up precisely to the observed change in the female participation rate due to second and higher order effects, we spread out the difference between the predicted and actual participation change in proportion to the first order effects. The unobserved age and cohort effects are in fact computed as composition effects. As the 2004 sample contains more younger cohorts than the 1992 sample, the younger cohort dummies receive a larger weight in 2004 and a positive overall cohort effect results.

Table 2.5.2: Decomposition of female participation growth 1992–2004

	Growth in participation rate (%-points)[a]	S.e.	Share of total growth (%)
Total growth 1992–2004	13.1		100
Household position	0.8	(0.0)	6
Having children	-0.1	(0.0)	-1
Having children × household position	0.0	(0.0)	0
Having children × education	0.3	(0.0)	2
Having children × year	5.3	(1.1)	40
Education	2.1	(0.3)	16
Education partner	0.3	(0.1)	2
Education × education partner	0.2	(0.1)	2
Unemployment	1.6	(0.5)	12
Unobserved age effects	-0.6	(0.4)	-4
Unobserved cohort effects	3.3	(1.5)	25

[a] Decomposition is based on marginal effects derived from parameter estimates for Model I (reported in tables 2.5.1 and 2.A.1). Second and higher order effects were spread out over all components according to shares reported in the last column.

The total change in the participation rate amounts to 13%-points over the period 1992–2004. It can be seen that nearly half of this effect has to do with the household situation. In particular, the participation decision has become less sensitive for raising children, i.e. women with children decide to participate in the labor market more often. On the other hand, the participation growth since 1992 is not related to women having less children. Another household effect predicted by the model results from the increasing share of single women and cohabiting women. According to the model, both these categories show higher participation rates than married females. The share of single women increases from 18% to 20% between 1992 and 2004, and that of cohabiting women from 9% to 15%, whereas the share of married women decreases from 63% to 57%.

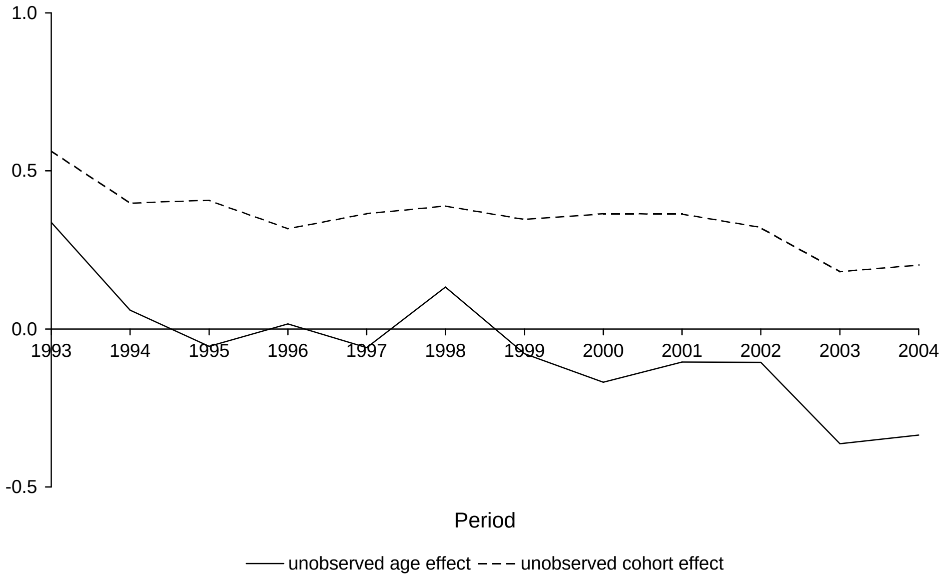
The second most important factor explaining the increasing participation rates are the unobserved cohort effects, which account for one quarter of total growth. During the period 1992–2004 many women from the pre-1955 generations retired from the labor market, while at the same time younger cohorts with relatively high unobserved cohort effects entered the labor market. According to our model, this composition effect led to an increase in the overall female participation rate by 3.3%-points.

Third, increasing levels of education account for about one sixth of total growth, and an encouraged worker effect of one eighth of the total growth is found. The encouraged worker effect results from the decrease in the delayed unemployment rate, which was 7.4% in 1992 and 5.7% in 2004. As a result, quite some women were attracted to the labor market (1.6%-point). Other effects, like the education of the partner, are less relevant in explaining participation growth.

In terms of statistical significance at the common 5% level, we may conclude that the demographic composition of households, education, changing effect of having children, unobserved cohort effects, and a cyclical encouraged worker effect have all contributed to the growth of the participation rate during the period.

Perhaps somewhat surprisingly, the age composition does not have a significant effect on the development of the participation rate. This may however change in the future, when the share of elderly in the population will increase further. To illustrate this, we have depicted the development of the age effect over time in figure 2.5. The trend is clearly negative, and will continue to be negative, so that statistical significance of a negative age effect appears to be just a matter of time. It also appears from the figure that the impact of the unobserved cohort effects decreases over time. This finding is particularly important for the future development of female labor force participation.

Figure 2.5: Age and cohort effects over time (in %-point)



Note: the lines show the effects of changing age and cohort compositions on the aggregate female participation rate.

2.6 Projected growth in two scenarios

In this section, we make a projection for the coming decades by substituting future projections for variables and unobserved age and cohort effects into Model I. Projections for all household variables and variables related to children are based on the long-term forecasts of Statistics Netherlands (Van Agtmaal-Wobma and van Duin, 2007). The demographic (long-term) projection is also taken from Statistics Netherlands (De Jong, 2005). Unemployment is fixed at the estimated equilibrium rate of 4.4% (CPB, 2006). An overview of the projections for the underlying variables is provided in table 2.6.1.

The fraction of women who live in a household without children increases by 4%-points. Furthermore, there will be more single women, more cohabiting women, and less married women. The fraction of higher educated women rises from 25 to 34%, and for male partners the same share of highly educated is achieved. This increase goes together with a decrease in the fractions of

Table 2.6.1: Assumptions made in scenarios

Variable(s)	Projection assumed
Household situation	Fraction of married women decreases to 0.42; fraction of cohabiting women who have (have not) been married increases to 0.04 (0.17); fraction of single women who have (have not) been married increases to 0.15 (0.15)
Having children	Fraction of women in a household without children increases slightly to 0.59, and consequently all categories of households with children show a small decrease.
Having children × year	No change (Basic scenario) or further participation increase among women with minor children (Emancipatory scenario): the post-2004 increase equals the estimated trend between 1998 and 2004 (see Table 2.A.1)
Education	Fraction with tertiairy education continues to grow to 0.34; fraction with primary (lower secondary) education continues to fall to 0.06 (0.20).
Unemployment	Convergence towards equilibrium unemployment rate (4.4%)
Unobserved age effects	No changes (only compositional effects)
Unobserved cohort effects	No changes compared to cohorts born after 1955 (see Figure 2.2)

primary and lower secondary educated individuals, with both groups showing a similar decrease by about 4%-points.

One of the most crucial parameters in the future projection is the development over time of the effect of having children. As the estimation results hardly provide any clue where this trend will stop, we have to make a rather arbitrary assumption on this. In order to investigate the impact of such an assumption, we construct two different scenarios. In the basic scenario we assume that there are no further developments in this parameter, i.e. the child effect remains constant at the 2004 level. In the alternative scenario, we assume that the remaining growth (after the year 2004) equals the growth during the period 1998–2004.[9]

[9] In the basic scenario the parameter for 'Having children under 18 × Year' is kept at the value of 0.78 (table 2.A.1). In the emancipatory scenario, this parameter is set at the value of 1.13 ($= 0.78 + (0.78 - 0.42)$).

An important argument for not postulating too large a growth is that child care facilities – especially government subsidies for families with children – are currently already at a high level (Jongen, 2010). While the introduction of these facilities has contributed to the past growth of the female participation rate, it seems unlikely that it will continue to do so. We postulate that the main driver of the increasing child effect is a more favorable attitude of women towards the combination of paid work and care for children. This may, for example, be the result of dynamics in these attitudes as nowadays non-employed women with children see more and more women with children that are employed.

Finally, we presume that both the unobserved cohort and age effects remain constant for future generations. The first was strongly suggested by the estimation results shown in figure 2.2, which shows that the unobserved cohort effect has been stable since the generation born around 1955. The second implies that unobserved age effects remain the same for future generations. A change in the participation rate for a given age is therefore assumed to occur only as a result of a change in exogeneous variables, such as the presence of children. There may still be an unobserved age effect at the aggregate level, due to a changing demographic composition. Likewise, an aggregate cohort effect is expected when the generations born before 1955 will be replaced by younger generations.

The resulting projected growth of the female participation rate is shown in table 2.6.2. The household situation, the fact that there will be less households with children, the increasing level of education, and a structurally lower unemployment rate all help to increase the participation rate from 2004 on. The effect of each of these factors is about 1.6%-points. These four factors are thus equally important for the future growth of female labor supply in the basic scenario.

The aggregate unobserved age effect is negative as the share of elderly women will rise. The aggregate unobserved cohort effect is positive since a part of the pre-1955 cohorts with relatively low participation rates will still be replaced by younger cohorts. As both effects have about the same size, they cancel out. In sum, the projected growth in the basic scenario amounts to 7%-points, which equals about half of the growth during the period 1992–2004.

In the 'emancipatory scenario', there is a large effect of changing attitudes towards the combination of paid work and family responsibilities, so that the projected participation growth equals nearly 10%-points.

Table 2.6.2: Decomposition of projected female participation growth 1992–2050 (%-points)

	2005–2050 sc. B[a]	sc. E[a]	1992–2004
Total growth	7.0	9.8	13.1
Household position	1.5	1.5	0.8
Having children	1.6	1.6	-0.1
Having children × household position	0.0	0.0	0.0
Having children × education	0.2	0.2	0.3
Having children × year	0.0	2.9	5.3
Education	1.6	1.6	2.1
Education partner	-0.1	-0.1	0.3
Education × education partner	0.4	0.4	0.2
Unemployment	1.6	1.6	1.6
Unobserved age effects	-3.1	-3.1	-0.6
Unobserved cohort effects	3.3	3.3	3.3

[a] B refers to the basic scenario, and E refers to the emancipatory scenario.

We have also checked the sensitivity of the basic scenario with respect to the education projection (not reported in the table). An important difference with the child/time-effect is that the education variables cause less uncertainty as a result of 'natural bounds' (e.g. it seems unlikely that the share of highly educated will exceed one half). For instance, if the growth in higher education participation is only half as great as described above, then total participation growth is lowered by just one %-point (in both scenarios).

A further increase in the female participation rate has important consequences for the government budget. Many industrialized countries have unsustainable government budgets given their aging populations. The main reason is that most social arrangements, like e.g. health care, social insurance, and pensions, are predominantly financed on a pay-as-you-go basis, whereas the elderly relatively often make use of these arrangements. An increasing female participation rate leads to higher employment, and this boosts economic growth. As a result, tax revenues will increase. To illustrate this for the Netherlands, an increase in the female labor participation rate by 7%-points as in the 'basic scenario' would improve the fiscal sustainability by about 1%-point of GDP.[10] In the 'emancipatory scenario' this would become 1.5%-points of GDP.

[10] A 2%-points increase in the labor force participation rate of women implies an improvement

Compared to the entire structural deficit in the Netherlands, this implies that about one third would be alleviated as a direct result of the increasing female participation rate.[11]

2.7 Conclusion and discussion

In the Netherlands, the female participation rate has increased considerably over the last decades. A further increase will help to alleviate the problem of fiscal sustainability due to the aging of the society. Which factors have played a role in the increase of the female participation rate between 1992 and 2004, and which further increase may we reasonably expect in the next decades? In order to answer these questions, we estimate a binary age-period-cohort model for the participation of women born between 1925 and 1986 on the basis of the Dutch Labor Force Survey 1992–2004. We use demographic forecasts of Statistics Netherlands to make an educated guess for the future development of female participation rates.

The estimation results indicate that the female participation rate is higher when the labor market is relatively tight. Between 1992 and 2004 the business cycle caused an increase in the female participation rate by nearly 2%-points. With regard to the household situation, single and cohabiting women have a relatively high probability to participate, while women with children have a relatively low probability to participate. The effect of having children is different for singles and for women part of a couple. Between 1992 and 2004 participation has become less sensitive for the presence of children, and this has played an important role in the increase of the participation rate. During the period under consideration, the availability and affordability of child care facilities improved substantially, and this may explain an important part of this effect. Furthermore, the increase in the education level between 1992 and 2004 has caused one sixth of the total increment in the female participation rate between 1992 and 2004.

of the fiscal sustainability by 0.3%-points of GDP, keeping other factors fixed (Van Ewijk et al., 2006).

[11] The sustainability gap is estimated at 3.4% of GDP keeping the female participation rate fixed (Van Ewijk et al., 2006).

Changes in unobserved cohort effects are very important for the cohorts born between 1935 and 1955. These effects account for about one quarter of the total increase in the female participation rate between 1992 and 2004. The estimated unobserved cohort effects are rather constant for the generations born after 1955. Note that the unobserved cohort effects should be interpreted with care as they may pick up time trends that we did not correct for. An obvious candidate for such a time trend would be the increase in real wages per level of education as it may have encouraged women to participate in the labor market. This may explain part of the increase in participation over the successive cohorts.

Although our research does not explicitly address the role of social norms and attitudes towards paid employment, it is conceivable they play a role in explaining the development of participation over successive cohorts. The reason is that sociological research on social norms and attitudes with respect to the combination of employment and family care responsibilities finds a development over the cohorts which is practically identical to the development we find for participation (SCP/CBS, 2006). The developments of participation over generations has been addressed in the international literature as well. Fernandez (2007) confronts a model of culture and intergenerational learning with the increase in US female participation. She concludes that culture can explain the particular pattern of the increase in participation. And there is evidence for the US that birth control has played a role as well. Goldin and Katz (2002) find that oral contraceptives, which became available in the years the generations born in the 1950s became mature, did affect family formation and the careers of women of this generation.

Using our estimation results, we have given some indication of the future prospects of female labor force participation. The female participation is likely to increase at a much lower rate, in particular as cohort effects have stabilized for the generations born after 1955, and we find no indications that social norms with regard to paid employment will evolve again. In the two scenarios we construct, the female participation rate increases by 7 and 10%-points, respectively, depending on the assumption made on the evolvement of attitudes towards the combination of paid work and children. A simple calculation shows that this increment alleviates the structural fiscal deficit caused by aging of the population by 1.0 to 1.5%-points of GDP.

2.A Extended estimation results

Table 2.A.1: Extended estimation results for table 2.5.1: education parameters and interaction effects

	Model I[a]		Model II[a]	
Variable	Coef.	S.e.	Coef.	S.e.
Child(ren) < 18 × 1992				
Child(ren) < 18 × 1993	0.10	(0.09)	0.13	(0.03)
Child(ren) < 18 × 1994[b]	-1.13	(0.09)	-1.10	(0.03)
Child(ren) < 18 × 1995	0.29	(0.10)	0.26	(0.03)
Child(ren) < 18 × 1996	0.30	(0.11)	0.28	(0.03)
Child(ren) < 18 × 1997	0.38	(0.11)	0.31	(0.03)
Child(ren) < 18 × 1998	0.42	(0.11)	0.46	(0.03)
Child(ren) < 18 × 1999	0.42	(0.11)	0.41	(0.03)
Child(ren) < 18 × 2000	0.39	(0.11)	0.44	(0.03)
Child(ren) < 18 × 2001	0.47	(0.12)	0.49	(0.03)
Child(ren) < 18 × 2002	0.59	(0.12)	0.56	(0.03)
Child(ren) < 18 × 2003	0.67	(0.13)	0.59	(0.03)
Child(ren) < 18 × 2004	0.78	(0.14)	0.60	(0.03)
Primary education (E1)				
Lower secondary education (E2)	0.55	(0.09)	0.78	(0.03)
Higher secondary education (E3)	0.73	(0.16)	1.09	(0.02)
Tertairy education (E4)	1.12	(0.14)	1.52	(0.03)
Partner: Primary education (EP1)[c]	-0.13	(0.06)	-0.14	(0.05)
Partner: Lower secondary educ. (EP2)[c]	0.13	(0.07)	0.12	(0.05)
Partner: Higher secondary educ. (EP3)[c]	0.30	(0.06)	0.29	(0.05)
Partner: Tertairy education (EP4)[c]	0.19	(0.06)	0.18	(0.07)

[a] Model I refers to the proxy variable approach, and Model II to the functional form approach.

[b] The interaction effect “Child(ren) < 18 × 1994” also contains the general effect of having at least one child younger than 18 in that year (see the three coefficients on “Age youngest child” in table 2.5.1). The reason is that the age of the youngest child in the year 1994 was not observed in the data set.

[c] It is possible to identify a full set of coefficients for the education level of the partner, because we can use single women as the reference category (for which all related dummy variables equal zero). As a consequence, the coefficients of four interaction dummy variables are not identified and need to be set to zero.

Table 2.A.1: Extended estimation results for table 2.5.1: education parameters and interaction effects *(continued)*

	Model I[a]		Model II[a]	
Variable	Coef.	S.e.	Coef.	S.e.
E1 × EP1[a]				
E1 × EP2[a]				
E1 × EP3[a]				
E1 × EP4[a]				
E2 × EP1	-0.26	(0.06)	-0.25	(0.04)
E2 × EP2	-0.37	(0.06)	-0.36	(0.03)
E2 × EP3	-0.37	(0.07)	-0.35	(0.03)
E2 × EP4	-0.28	(0.06)	-0.29	(0.06)
E3 × EP1	0.12	(0.09)	0.16	(0.04)
E3 × EP2	-0.04	(0.09)	0.00	(0.03)
E3 × EP3	-0.06	(0.09)	-0.02	(0.03)
E3 × EP4	-0.09	(0.08)	-0.07	(0.06)
E4 × EP1	-0.04	(0.13)	-0.03	(0.09)
E4 × EP2	0.04	(0.08)	0.06	(0.05)
E4 × EP3	-0.05	(0.07)	-0.03	(0.04)
E4 × EP4	0.02	(0.06)	0.03	(0.06)

[a] These dummy coefficients are set to zero for identification; see note c in the previous table.

Table 2.A.2: Extended estimation results for table 2.5.1, model II (functional form approach): year and age dummy coefficients

Var.	Coef.	S.e.	Var.	Coef.	S.e.	Var.	Coef.	S.e.
1992	0.00	-	age 26	2.73	(0.04)	age 47	2.70	(0.06)
1993	-0.05	(0.02)	age 27	2.80	(0.04)	age 48	2.61	(0.06)
1994	-0.10	(0.02)	age 28	2.86	(0.04)	age 49	2.48	(0.06)
1995	-0.12	(0.02)	age 29	2.89	(0.04)	age 50	2.35	(0.07)
1996	-0.14	(0.03)	age 30	2.92	(0.04)	age 51	2.26	(0.07)
1997	-0.10	(0.03)	age 31	2.91	(0.04)	age 52	2.20	(0.07)
1998	-0.18	(0.03)	age 32	2.95	(0.04)	age 53	2.12	(0.07)
1999	-0.08	(0.03)	age 33	2.98	(0.04)	age 54	2.00	(0.08)
2000	-0.11	(0.03)	age 34	2.97	(0.04)	age 55	1.87	(0.08)
2001	-0.07	(0.03)	age 35	3.03	(0.04)	age 56	1.78	(0.08)
2002	-0.04	(0.03)	age 36	3.07	(0.04)	age 57	1.57	(0.08)
2003	-0.04	(0.04)	age 37	3.06	(0.04)	age 58	1.44	(0.09)
2004	-0.04	(0.04)	age 38	3.09	(0.05)	age 59	1.18	(0.09)
age 18	0.00	-	age 39	3.11	(0.05)	age 60	0.79	(0.10)
age 19	0.37	(0.03)	age 40	3.11	(0.05)	age 61	0.35	(0.11)
age 20	0.80	(0.03)	age 41	3.07	(0.05)	age 62	0.16	(0.11)
age 21	1.17	(0.03)	age 42	3.06	(0.05)	age 63	-0.10	(0.12)
age 22	1.57	(0.04)	age 43	3.00	(0.05)	age 64	-0.18	(0.13)
age 23	1.92	(0.04)	age 44	2.93	(0.05)	age 65	-0.69	(0.16)
age 24	2.27	(0.04)	age 45	2.90	(0.06)	age 66	-0.49	(0.17)
age 25	2.57	(0.04)	age 46	2.79	(0.06)	age 67	-0.52	(0.18)

3 Pathways to Retirement and Cause-Specific Mortality Risks in the Netherlands

This chapter is based on Kalwij, Alessie, and Knoef (2010).

3.1 Introduction

The aging of the Dutch population has raised concerns about the sustainability of the welfare state as it increases public expenditures on, for instance, long-term care and retirement pensions (Van Ewijk et al., 2006). One means of alleviating the burden of an aging population on public finances is to increase labor force participation and so raise social security contributions and tax revenues. Since the 1990s, social security programs and pension schemes are, therefore, being redesigned to create stronger incentives for continued work at older ages. These reforms, together with the increased labor force participation of women, are likely to have contributed to the rising participation in the workplace of the 55–64 population from under 30% in the mid-1990s to 45% in 2007 (Euwals et al., 2009, Van Oorschot, 2007). Recent reforms like the 2004 introduction of job search requirements for older unemployed persons (De Vos et al., 2010), the abolishment of the favorable fiscal treatment of early retirement contributions (CPB, 2005) and a tax exemption for individuals who continue working after age 62 (Stimulansz, 2009), as well as the proposed rise in the statutory retirement age from 65 to 67 (CPB, 2009 and 2010), are expected to further increase employment among the 55+ population.

Nevertheless, the success of, and the political support for, policies aimed at keeping older workers in employment depends, among other things, on the health conditions of these workers. One particular concern is that workers who leave employment before the statutory retirement age of 65 may have health conditions that prevent them from remaining employed until age 65. An obvious group of such older workers is those drawing disability insurance benefits, but the concern extends to other groups like early retirees who for health reasons may have chosen to leave the labor force before age 65. Although several empirical studies have addressed this issue by estimating the impact of health on employment, the findings appear to depend, among other things, on the available health measures such as self-reported health status, (objective) health conditions or (future) mortality risk. For instance, for the Netherlands Kerkhofs et al. (1999) find no effect of an objective health indicator on the transition from employment to early retirement while Lindeboom and Kerkhofs (2009) find that individuals in bad health are more likely to leave employment to early retirement when using a self-reported work-related health measure. The findings for the U.S. in Bound et al. (1999) and for Canada in Campolieti (2002) are in line with this latter result. These studies often employ an instrumental variables approach to take into account that, for reasons such as measurement error and simultaneity, the health measure is likely to be an endogenous explanatory variable in an employment equation (e.g., Stern, 1989).[1]

An alternative approach to assess whether individuals who leave employment early may have health conditions that prevent them from remaining employed, and the one taken in this study, is to estimate the association between the different labor market states before retirement and mortality risk during retirement. A few empirical studies have taken this approach and the findings are contradictory. For instance, Tsai et al. (2005) and Bamia et al. (2007) find an increased mortality risk among early retirees in the U.S. and Greece, respectively, whereas Brockman et al. (2009) and Litwin (2007) find no such increased risk among early retirees in Germany and Israel, respectively. Likewise, Iversen et al. (1987) have shown an increased mortality risk among unemployed individuals. One reason to expect a priori that older workers who leave employment to early retirement or unemployment have, on average, no increased mortality risk in countries like the Netherlands, is the presence of a disability insurance scheme that selects individuals with health conditions

[1]See, for instance, Bound (1991) and Baker et al. (2004) for a critical discussion of this approach.

that limit their work capacity out of the labor force. This selection implies an increased mortality risk among individuals drawing disability insurance benefits. Empirical support for this implication is available for Germany (Brockman et al., 2009), Norway (Gjesdal et al., 2007), and Sweden (Karlsson et al., 2007, and Wallman et al., 2006).

This study contributes to the literature on the health conditions of the unused labor capacity among the 58–64 population in the Netherlands by taking the alternative approach discussed above and using a large individual-level administrative dataset that contains information on labor market status before retirement and mortality during retirement. Accordingly, we estimate the association between the pathways to statutory retirement at age 65 and cause-specific mortality risk after the statutory retirement age of 65. The pathways to retirement delineated here are the years of being employed, self-employed, unemployed, nonparticipating, early retired, or on disability insurance benefits between the ages of 58 and 64. This latter pathway concerns individuals who leave employment early because of health conditions that limit their work capacity. We relate these pathways to three competing causes of death – cancer, cardiovascular disease (CVD), and other diseases. Differentiating these three causes of death may yield additional insights as, for instance, early retirement is often related to CVD mortality risk (e.g., Bamia et al., 2007).

Based on estimated associations we draw conclusions concerning the health conditions that may limit the work capacity of older workers who leave employment early and this requires three assumptions. First, we assume, in line with Grossman's health stock model (Grossman, 2000), that pre-retirement health conditions are related to later life health and mortality risk. Empirical support for this is provided in, for instance, Portrait et al. (2001). Second, we assume a positive association between pre-retirement health conditions and health conditions that limit a person's work capacity.[2] Lindeboom and Kerkhofs (2009), for instance, provide empirical support for this assumption. Third, and perhaps most importantly, we assume that an early withdrawal from the labor market has no health-preserving effect. If individuals who leave employment improve their health relatively to individuals who remain employed then mortality risk during retirement overestimates their health status at the time they left employ-

[2]This does not mean we exclude the possibility that some pre-retirement health conditions that limit a person's work capacity do not affect mortality risk.

ment. Recent empirical evidence that uses (exogenous) variation in retirement policy to identify a possible causal effect of early retirement on mortality risk, however, suggests the contrary. For instance, Coe and Lindeboom (2008) find no causal effect of early retirement on mortality risk in the U.S., Kuhn et al. (2010) find a positive causal effect of early retirement on CVD mortality risk in Austria for men (but not for women) and Coe and Zamarro (2008) report for Europe a health-preserving effect of retiring at age 65 but not of retiring at younger ages. The findings of Snyder and Evans (2006) for the U.S. suggest that post-retirement part-time work may have a health-preserving effect. In addition, it has been shown that early retirement has a negative causal impact on cognitive health (Rohwedder and Willis, 2010, and Bonsang et al., 2010). Likewise, Sullivan and Von Wachter (2009) and Eliason and Storrie (2009) show that job displacement increases mortality risk.

The recent empirical literature supports the above mentioned assumptions. This implies that if, for example, we would find no increased mortality risk after age 65 among early retirees, that they have, on average, no worse health and reduced work capacity than individuals who remained employed. Alternatively, in this example, if we would find an increased mortality risk after age 65 among early retirees then it can be that the relatively unhealthy workers (with reduced work capacity) retire early or that early retirement causes a higher mortality risk.

Our methodological framework for analyzing the associations between the pathways to retirement and cause-specific mortality risk is a dependent competing mortality risks model. The model allows for dependencies between the different competing risks by conditioning on both observed characteristics and an unobserved individual specific characteristic (see, e.g., Katsahian et al., 2006).[3] The model therefore takes into account that with age the sample becomes more selective in terms of both the observed and unobserved characteristics (e.g., Van den Berg, 2001). In addition we use information on multiple causes of death to allow the impact of the unobserved individual specific characteristic to differ across the cause-specific mortality risks.

The chapter is organized as follows. Section 3.2 describes the data. Section 3.3 outlines the statistical model, and section 3.4 reports the empirical results. Section 3.5 summarizes the main findings and concludes the chapter.

[3]For example, an unobserved individual specific characteristic such as smoking behavior affects mortality risks for lung cancer as well as for cardiovascular diseases.

Data 3.2

The data are taken from the 1989–2007 Income Panel Study of the Netherlands (IPO, Inkomens Panel Onderzoek, CBS 2009a) and the 1997–2008 Causes of Death registry (DO, Doodsoorzaken, CBS 2009b), both gathered by Statistics Netherlands. The IPO, a representative sample of the Dutch population, consists of an administrative panel dataset of, on average, about 95,000 selected individuals per year who are followed longitudinally. Sampling is based on individuals' national security number, and the selected individuals are followed for as long as they are residing in the Netherlands on December 31 of the sample year. Individuals born in the Netherlands enter the panel for the first time in the year of their birth, and immigrants to the Netherlands in the year of their arrival. The main advantages of using this administrative dataset compared to using survey data for our analysis are, apart from the large sample size, twofold. First, the IPO dataset includes individuals living in institutions for the elderly, such as nursing homes, who are usually absent or underrepresented in household surveys. Second, an individual only exits the panel on death or emigration from the Netherlands. To summarize, the IPO sampling framework guarantees a representative sample of the population and, therefore, our analysis is not compromised by possible sample selection issues that may be related to individuals' health status or mortality risk.

The IPO contains data on gender, age, marital status, income, homeownership, and labor market status. These data are obtained from official institutions, most particularly, the population registry and the tax office. The DO, on the other hand, provides information on the date of death and at most four causes of death for all residents deceased during the 1997–2008 period. Multiple causes of death include a primary cause and up to three secondary causes. These data, all registered using version 10 of the International Classification of Diseases (ICD10),[4] are provided by medical examiners who are legally obliged to submit them to Statistics Netherlands. The DO dataset also assigns a personal identifier that allows determination of whether an individual in the IPO has died by the next calendar year and, if so, the causes of death.

We select individuals who turned 65 during the 1996–2007 period, meaning that individuals from the oldest cohort (born in 1931) are included in the

[4]http://www.who.int/classifications/apps/icd/icd10online

data for at most 12 years and individuals from the youngest cohort (born in 1942) are included for only 1 year. This selected sample consists of 57,757 observations for 10,013 individuals. We remove 3.8% of the observations because of negative or zero income or missing values on the variables included in the empirical analysis. The hypothesis of the same mortality rate among the individuals excluded and those included is not rejected,[5] which implies no endogenous sample selection. Panel attrition for reasons other than mortality (mainly emigration) is about 0.12% per year. The resulting sample consists of 55,553 observations for 9,618 individuals, 1,147 of whom died during the sample period.

3.2.1 Variable definitions and descriptive statistics

We define age as the individual's age on January 1 of each year because in the Netherlands, the calendar year is also the fiscal year for income measurement, meaning that this choice ensures that income at age 65 is measured over the first entire calendar year of retirement. The dependent variable in our analysis is cause-specific mortality after the statutory retirement age of 65. For this variable, we distinguish between the two major causes of death, cancer (ICD10 code C00-D48) and CVD (ICD10 code I00-I99), and refer to all other causes of death as "other diseases". These latter, which include infectious diseases, diabetes, pneumonia, diseases of the digestive system, and fatal injuries, cannot be further disaggregated because small numbers of persons in our sample die from each of these diseases. The explanatory variables are gender, marital status at age 65, homeownership at age 65, standardized household income at age 65, and labor market status from age 58 onward. Table 3.2.1 describes how these variables, defined below, relate to cause-specific mortality rates.

Table 3.2.1 confirms the accepted pattern that men have a higher mortality rate than women: the three rightmost columns show that differential mortality with respect to gender appears to be equally strong for the different causes of death. The marital status variable distinguishes between widowed individuals, other single adult households including divorcees (hereafter, single), and married or cohabiting couples (married). As table 3.2.1 shows, individuals

[5] The corresponding p-value is 0.214.

Table 3.2.1: Cause-specific mortality rates by socioeconomic group

		Cause-specific mortality rate[a]			
	Sample proportion	All causes	Cancer	CVD[b]	Other diseases
	%	%	%	%	%
All	100	2.1	0.9	0.6	0.6
Gender					
Men	48.4	2.7	1.2	0.8	0.7
Women	51.6	1.5	0.7	0.4	0.4
Marital status at age 65					
Single	12.7	2.8	0.9	1.0	0.9
Widowed	11.9	2.4	1.0	0.6	0.8
Married	75.4	1.9	0.9	0.5	0.5
Birth cohort					
1931–1935	38.9	2.4	1.0	0.7	0.6
1936–1942	61.1	1.6	0.8	0.4	0.4
Homeowner at age 65					
No	49.9	2.5	1.0	0.7	0.8
Yes	50.2	1.6	0.8	0.5	0.3
Labor market status at age 58					
Employed	29.3	1.9	1.0	0.5	0.4
Self-employed	7.6	2.2	1.5	0.4	0.3
Unemployed	11.2	2.4	0.9	0.9	0.7
On disability	13.1	3.2	1.1	1.0	1.1
Early retired	14.4	2.0	0.9	0.5	0.6
Nonparticipating	24.4	1.5	0.7	0.4	0.4
Labor market status at age 62					
Employed	10.6	2.0	0.9	0.6	0.5
Self-employed	5.8	2.1	1.4	0.4	0.3
Unemployed	8.9	2.3	0.9	0.8	0.7
On disability	14.3	3.3	1.2	1.1	1.0
Early retired	36.6	2.0	1.0	0.5	0.5
Nonparticipating	23.7	1.4	0.6	0.4	0.4
Standardized household income	Sample mean (euros)				
1st quartile	11,208	3.0	1.1	1.0	0.8
2nd quartile	15,438	2.0	0.9	0.5	0.6
3rd quartile	19,251	1.7	0.9	0.4	0.4
4th quartile	32,320	1.6	0.8	0.4	0.4

[a] The mortality rate is defined as the probability of death within one year (in %).
[b] CVD = cardiovascular diseases

married at age 65 have relatively lower mortality rates for CVD and other diseases but about the same cancer mortality rate as individuals who are single or widowed at age 65. Homeowners at age 65 have a lower mortality rate than renters, an observation particularly related to the mortality risks for other diseases, but also for CVD.

The IPO income data, which are based primarily on records from the tax office and institutions that pay out (insurance) benefits, contain detailed and accurate information on all income components on an individual level. The income components are the total yearly amounts received from the different income sources. Based on the largest income component, Statistics Netherlands assigns one of the following labor market statuses to each individual: employed, self-employed, unemployed (receiving unemployment insurance or assistance benefits), disability (receiving disability or (long-term) sickness insurance benefits),[6] early retired (receiving pension income before the age of 65), or nonparticipating (receiving no labor income, pension, or benefits). Eligibility for disability insurance benefits is assessed by a medical doctor based on health conditions that adversely affect individual work capacity.[7] We separately distinguish the self-employed since they have an own responsibility to take out, for instance, disability insurance and this may affect the possible pathways to retirement.

Although our selected sample starts in 1996, the IPO dataset also contains information on the labor market status for all individuals in our sample from 1989 onward, meaning that we observe labor market status for all these individuals from age 58 onward. As shown in column 1, table 3.2.1, the 36.9% employment rate for 58-year-old individuals drops to 16.4% by age 62, while the proportion of those on disability increases slightly from 13.1% at age 58 to 14.3% at age 62. Between ages 58 and 62, the percentage of early retirees increases from 14.4% to 36.6%. We also observe a higher mortality rate related

[6]The disability insurance scheme is referred to as the WAO or WIA and the (long-term) sickness insurance scheme as the ZW (see www.uwv.nl). If, for example, an individual is partially disabled but the main source of income is employment, we classify this person as employed.

[7]Since the early 1990s, the disability insurance system has been reformed toward stricter eligibility conditions (see, e.g., De Vos et al., 2010), which may have caused individuals who begin claiming disability insurance benefits to become, on average, less healthy over time. However, in our empirical analysis we found no evidence for this assumption when testing if the impact of the pathway "years on disability" on mortality during retirement varies with the year when the claim started.

primarily to CVD and other diseases among individuals on disability but a higher mortality rate from cancer among the self-employed. Apart from these observations, no clear pattern emerges on cause-specific mortality rates by labor market status at ages 58 and 62.

For both the individual and the spouse (when present), we use income at the age of 65 as a proxy for lifetime income, which the health-economic theory suggests is associated with mortality risk (see, e.g., Grossman, 2000). Standardized household income is defined as an individual's lifetime income together with that of the spouse when present, gross of taxes and social insurance contributions, measured in 2005 euros using the consumer price index and, for married individuals, divided by the equivalence scale provided by Statistics Netherlands (Siermann et al., 2004).[8]

As table 3.2.1 shows, in line with the findings of other European studies (see, e.g., Kalwij et al. 2009, for a summary), the mortality rate is about twice as high among individuals in the lowest quartile of the income distribution than among individuals in the highest quartile (3.0% versus 1.6%). Income also appears most strongly associated with CVD mortality and most weakly linked to cancer mortality. These findings conform, to those of, for instance, Huisman et al. (2005) who, using data from eight European countries, report higher mortality risk among low- than among highly educated groups for all causes of death other than prostate cancer for men and lung cancer for women.

Although the cause-specific mortality rates given in table 3.2.1 refer to the so-called primary cause of death, medical examiners also report up to three contributing causes often termed secondary causes of death (see table 3.2.2). For example, if an individual with a fatal form of cancer dies from CVD, the medical examiner reports CVD as the primary cause of death and cancer as the secondary cause (as in 11 cases listed in table 3.2.2). Nevertheless, given the level of aggregation, the secondary cause of death may be the same as the primary one; for instance, an individual may suffer from two different cancers, making cancer both the primary and secondary cause of death (as in 22 cases given in table 3.2.2). About one-third of the recorded deaths in

[8]We exclude income from other household members, present in about 12% of the households but mostly children. This exclusion does not affect our results. Although income includes pension, labor, transfer, and capital income for individuals 65 and older, over 90% of income consists of public retirement pension benefits (independent of past earnings) and occupational pension income (see Knoef et al., 2009, for further details on the income measures in the IPO).

our sample have a secondary cause of death that is different from the primary cause. The next section discusses how we use this additional information to allow an unobserved individual specific characteristic to have different impacts on the cause-specific mortality risks. For a detailed discussion of the multiple causes of death and Statistics Netherlands' method for recording them we refer to Mackenbach et al. (1995, 1997).

Table 3.2.2: Multiple causes of death

Cell: number of deaths		**Secondary cause of death**		
Primary cause of death		Cancer	CVD[a]	Other diseases
Cancer	512	22	56	135
CVD	329	11	73	112
Other diseases	306	9	68	158
All causes	1,147	42	197	405

[a] CVD = cardiovascular diseases

3.3 Cause-specific mortality model

Our empirical model for analyzing cause-specific mortality risk is a discrete-time competing risks model that allows for dependency across these risks through time-invariant observed and unobserved heterogeneity. The dependent variable is whether an individual is deceased by the next calendar year and if so, whether the cause of death is cancer, CVD, or other diseases.

The cause-specific mortality risk conditional on age and individual characteristics is formalized as follows

$$\begin{aligned} &Pr\big(M_{j,a+1}(i)=1 \mid M_{j,a}(i)=0, X_a(i)=x_a(i), \eta(i); \beta_j, \alpha_j\big) \\ &\quad = F_j\big(x_a(i)\beta_j + \alpha_j \eta(i)\big), \end{aligned} \tag{3.1}$$

where $F_j(.)$ is the logistic cumulative distribution function that corresponds to mortality cause j ($j \in \{1,2,3\}$). $M_{j,a+1}(i)$ is equal to 1 if individual i became a years old and died at age $(a+1)$ from cause j, and 0 otherwise. $x_a(i)$ is a $(1 \times k)$ vector of an individual's observed characteristics at age a with a corresponding $(k \times 1)$ parameter vector β_j.

The time-invariant unobserved individual characteristic (a random effect) is denoted by $\eta(i)$ and is assumed to be normally distributed and independent of individual's observed characteristics at age 65. The α_j parameters allow the unobserved individual specific characteristic to have different impacts on the cause-specific mortality risks. As mentioned in the introduction, by modeling unobserved individual characteristics the model takes into account that with age the sample becomes more selective in terms of not only observed but also unobserved characteristics (e.g., Van den Berg, 2001).

The age dependency of cause-specific mortality risk is modeled using a linear age function (in the index).[9] We include the years in each labor market state from age 59 until age 64 to measure the extent to which pathways to retirement other than employment are associated with cause-specific mortality risk during retirement.

We estimate the model with two empirical specifications. The first specification only controls for gender, birth cohort, pathways and labor market status at age 58 and does not control for socioeconomic status. It also does not control for the time-invariant unobserved individual characteristics. The inclusion of labor market status at age 58 ensures that the associations between the pathways and cause-specific mortality risk are identified solely from labor market transitions after age 58. The estimated associations provide, under the assumptions that we discussed in the introduction, insights into, for example, whether individuals who leave employment between the ages of 58 and 65 to early retirement or unemployment, may have health conditions that prevent them from remaining employed.

The second specification controls for gender, birth cohort, pathways to retirement, labor market status at age 58 and socioeconomic status using the characteristics marital status, homeownership and the logarithm of standardized household income. In addition, it controls for time-invariant unobserved individual characteristics. Socioeconomic status is known to be strongly associated with mortality risk and this second specification provides further empirical evidence on these associations for the Netherlands.[10] Also, conditioning on

[9] Kalwij et al. (2009) present empirical evidence in favor of using a linear age function.

[10] We refer, for instance, to Duleep (1986), Huisman et al. (2004, 2005), Lindahl (2005), Marmot et al. (1991), Menchik (1993), Smith (1999) and Van Kippersluis et al. (2010) for empirical evidence and discussions on the socioeconomic variation in health- (and cause-) specific mortality risk(s). Grundy and Holt (2001) argue that homeownership is a good proxy

(observed) socioeconomic status provides insights into the extent to which workers with different pathways to retirement differ with respect to socioeconomic status and corresponding health status. Should the selection into a specific pathway be health related then we expect the associations between the pathways and mortality risk to be affected by the inclusion of socioeconomic status variables that are, in turn, known to be strongly related to mortality risk. If, for instance, we find that by controlling for socioeconomic status the (possible) positive association between a pathway (relative to continuing employment) and mortality risk diminishes, this would provide empirical evidence that workers with relatively lower socioeconomic status (and corresponding lower health status) are more likely to take this pathway. In addition, by including socioeconomic status variables measured at age 65 we take into account the possible impact on mortality risk of changes in these variables that may result from leaving employment before age 65.[11]

3.3.1 Model estimation

We observe an individual from age 65 until death or until the last sample year. With $A(i)$ denoting the age of the individual when last observed in the sample, the variable $m_j(i)$ is equal to 1 if j is the cause of death at age $A(i)+1$ ($j \in \{1,2,3\}$), and 0 otherwise. We summarize all information observed for individual i in the vector $z(i) = \{x_a(i)\}_{a=65}^{A(i)}$. The probability of individual i surviving from age 65 to age $A(i)$ and either surviving to or being deceased at age $A(i)+1$ from mortality cause j is given by

$$
\begin{aligned}
&P\left(m_1(i), m_2(i), m_3(i) \mid z(i), \eta(i); \beta, \alpha\right) \\
&= \prod_{j=1}^{3} \left(\left(\prod_{a=65}^{A(i)-1} \left(1 - F_j\left(x_a(i)\beta_j + \alpha_j \eta(i)\right)\right)\right)^{I(A(i)>65)} \right. \\
&\qquad \left. \times \left(1 - F_j\left(x_{A(i)}(i)\beta_j + \alpha_j \eta(i)\right)\right)^{(1-m_j(i))} \right.
\end{aligned}
$$

for older people's socioeconomic status.

[11] There are, however, few changes in socioeconomic status variables between the ages 58 and 65. Also (or instead) including, for instance, marital status and homeownership at age 58 does not affect the main conclusions of this study.

$$\times \left(F_j \left(x_{A(i)}(i)\beta_j + \alpha_j \eta(i) \right) \right)^{m_j(i)} \Bigg), \tag{3.2}$$

where $\beta = (\beta_1, \beta_2, \beta_3)$ and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. The first term in the right-hand side of equation (3.2), in between the square brackets, is the survival probability up to $A(i)$, the second term is the probability of surviving one more year, and the third term is the probability of being deceased from cause j at age $A(i) + 1$.

As previously explained, $\eta(i)$'s are unobserved random effects, meaning there is no empirical counterpart to equation (3.2). We therefore assume that the random effects are normally distributed with a zero mean and a variance of σ^2 and take the conditional expectation of equation (3.2) with respect to η:

$$\begin{aligned} &E_\eta \left(P(m_1(i), m_2(i), m_3(i) \mid z(i), \eta(i); \beta, \alpha) \right) = \\ &\int_{-\infty}^{\infty} P\left(m_1(i), m_2(i), m_3(i) \mid z(i), \eta(i); \beta, \alpha\right) d\Phi \left(\frac{\eta(i)}{\sigma} \right), \end{aligned} \tag{3.3}$$

where $\Phi(.)$ is the cumulative normal distribution function. The maximum likelihood estimates of the model parameters are given by

$$\left(\widehat{\alpha},\ \widehat{\beta},\ \widehat{\sigma}\right) = \underset{\alpha,\beta,\sigma}{\operatorname{argmax}} \sum_{i=1}^{n} \log E_\eta \left(P(m_1(i), m_2(i), m_3(i) \mid z(i), \eta(i); \beta, \alpha) \right) \tag{3.4}$$

where n is the number of individuals. We evaluate the integral of equation (3.4) using a Gaussian quadrature (see, e.g., Cameron and Trivedi, 2005).[12]

Model identification 3.3.2

A well-known feature of a competing risks model is that one has to make an assumption concerning the dependency across the different risks in order to identify the parameters (see, e.g., Van den Berg, 2005). Frequently, such risks are assumed to be independent conditional on the observed covariates (see, e.g, Yashin et al., 1986). Yet in an analysis of cause-specific mortality risk, this latter assumption may be unrealistic (for a discussion on this, see, e.g., Mackenbach et al., 1995 and Vaupel and Yashin, 1999). We therefore allow for dependency between cause-specific mortality risks by including an individual specific

[12]For estimation, we use the MATA module of the STATA software (www.stata.com).

random effect as formulated in equation (3.1). A more general specification, such as different random effects for each risk, is impossible because of data limitations: it would require, among other things, at least three continuous covariates. However, one important aspect of the model given in equation (3.1) is that, in the second empirical specification, it contains one continuous covariate (standardized household income) which, together with the proportionality between age pattern, covariates, and the random effect imposed on mortality risk, ensures identification of the random effect distribution. A formal discussion of the identifiability of mixed proportional hazards in competing risks models can be found in, for instance, Heckman and Honoré (1989) and Abbring and van den Berg (2003).

Our model also allows the (time-invariant) unobserved individual specific characteristic to have different impacts on the cause-specific mortality risks through the α parameters. Only their relative sizes are identified and we normalize $\alpha_1 = 3 - \alpha_2 - \alpha_3$.[13] The identification of the α parameters in equation (3.1) is guaranteed by the fact that medical examiners report more than one cause of death for about one-third of deaths (see table 3.2.2). As regards the likelihood function in equation (3.2), this information on multiple causes of death means that $m_j(i)$ can be equal to 1 for more than one j for individual i. Among others, Israel et al. (1986) and Mackenbach et al. (1999) suggest to use multiple causes of death statistics to obtain information about associations between different cause-specific mortality risks. Multiple causes of death statistics have, as far as we know, not previously been used in the context of a dependent competing risk model that includes unobserved individual specific characteristics (random effects). Instead, for instance, Mackenbach et al. (1999) use a logistic regression model to estimate the prevalence of a specific secondary competing cause of death, given the primary cause of death. Finally, as discussed above, the model includes a separate control for labor market status at age 58 to ensure that the associations between the pathways to retirement (measured by the years in the different labor market statuses from age 59 onward) and cause-specific mortality risk are identified solely from labor market transitions after age 58.

[13]This specification allows one or two of the α_j parameters to be equal to zero.

Estimation results 3.4

The estimation results from the model outlined above are given in tables 3.4.1 and 3.4.2. In the following discussion we use a 5% level of significance. Table 3.4.1 shows the results using the first empirical specification, i.e. without controlling for socioeconomic status and unobserved individual specific characteristics. The table shows that women have a lower mortality risk than men for all three causes of death, and that individuals in the more recent birth cohorts have a lower mortality risk from CVD.[14] Being on disability is the only pathway to retirement that is significantly associated with mortality risk during retirement for all three causes of death.

Table 3.4.2 shows the results of the second empirical specification which controls for observed socioeconomic status and unobserved individual specific characteristics. For all three causes of death, mortality risk is higher among singles than married individuals. Homeowners have lower mortality risks, especially from other diseases, and household income is more strongly and negatively associated with the mortality risks from CVD and other diseases than with that from cancer. This latter finding is in line with results for the U.S. (e.g., Cutler et al., 2006, and Smith, 1999) and the U.K. (e.g., Marmot et al., 1991). Finally, being on disability is the only pathway that is significantly associated with mortality risk during retirement and for all three causes of death. This latter finding is qualitatively similar to that in table 3.4.1. The significant and relatively large estimated standard deviation of the random effect (at the bottom of the table) suggests an important role for unobserved individual-specific characteristics.[15] The hypothesis that α is equal to zero is rejected for each of the three risks (based on t-statistics) and, more importantly, we reject the hypothesis of the same (relative) impacts of unobserved individual specific characteristics on all three mortality risks.[16] These results on the standard deviation of the random effect and the α parameters imply dependence between the three competing risks, conditional on observed individual characteristics.

[14]In line with this finding, Janssen et al. (2004) show that old-age mortality from cardiovascular diseases has declined after the 1960s.

[15]The parameter estimates are affected by excluding random effects. For instance, and in line with the results in tables 3.4.1 and 3.4.2, it yields an underestimated age effect. This may reflect that with age, the population at risk consists of relatively fewer frail individuals.

[16]The p-value corresponding to $H_0 : \alpha_2 = \alpha_3 = 1$ is equal to 0.034 (note that $\alpha_1 = 3 - \alpha_2 - \alpha_3$).

Table 3.4.1: Estimation results (excluding the controls for socioeconomic status)

Dependent variable: Cause-specific mortality risk	**Cancer**		**CVD**		**Other diseases**	
Covariate[a]	Parameter estimate	Standard error	Parameter estimate	Standard error	Parameter estimate	Standard error
Age	0.045	(0.019)	0.021	(0.020)	0.073	(0.019)
Birth cohort	-0.016	(0.019)	-0.103	(0.021)	-0.055	(0.020)
Woman	-0.594	(0.121)	-0.735	(0.139)	-0.609	(0.129)
Years self-employed between ages 58 and 64	0.045	(0.091)	-0.061	(0.129)	0.028	(0.121)
Years unemployed between ages 58 and 64	0.041	(0.065)	-0.023	(0.079)	-0.004	(0.071)
Years on disability between ages 58 and 64	0.185	(0.069)	0.156	(0.076)	0.195	(0.068)
Years nonparticipating between ages 58 and 64	0.023	(0.063)	-0.108	(0.079)	-0.095	(0.075)
Years in early retirement between ages 58 and 64	0.045	(0.045)	-0.058	(0.053)	-0.019	(0.052)
Log-likelihood value	-8562.7					
Number of individuals	9618					

[a] The model also includes a constant and dummy variables for the different labor market statuses at age 58 (see section 3.3).

Nevertheless, although the parameter estimates in table 3.4.2 provide insights into the direction and relative size of the associations between the covariates and the cause-specific mortality risks, they offer no clear quantitative insights. We therefore use these estimates to predict the cause-specific and all-causes mortality probabilities of a reference individual, conditional on surviving up to age 65, with specific values assigned to the covariates.[17] Here, we take as a reference a male born in 1931, who is married at age 65, has a median standardized household income, lives in a rented accommodation, and who stayed employed up to the statutory retirement age of 65.[18] First, we predict the cause-specific and all-causes probabilities of the reference individual having died by ages 70, 75, and 80 (see table 3.4.3); next, we change his characteristics one at a time and report the resulting changes in these probabilities of his having died by age 75 (see table 3.4.4).

As table 3.4.3 shows, with age there is a strong increase in the probability of the man having died and a decrease in the relative importance of cancer as the cause of death. Specifically, cancer accounts for about half of deaths by age 70 but only about one-third by age 80. The relative importance of CVD, however, remains fairly constant with age, and the relative importance of other diseases rises. As the first line of table 3.4.4 (like that of table 3.4.3) shows, by age 75, the reference male has a 13% probability of having died and this can be disentangled in a 5% probability of having died from cancer, a 4% probability of having died from CVD, and a 4% probability of having died from another disease.

In contrast, a woman with the same characteristics as the reference man has about a 9 percentage point lower probability of having died by age 75 (table 3.4.4). The reduction in mortality is of similar size for all three causes of death. Moreover, compared to the reference individual, who is married at age 65, a single individual has a two times higher probability of having died by age 75, an increase related primarily to CVD and other diseases. Individuals that are widowed at age 65 have about a 9 percentage point higher probability of having died by age 75, primarily related to other diseases. The younger birth cohorts, however, have a lower probability

[17]These calculations are based on Monte Carlo simulations. In a competing risk setting in discrete time, two events can occur at the same time; however, by using random assignment in the simulation, we allow for only one cause of death (cancer, CVD, or another disease).

[18]The choice of the reference individual only affects the baseline predictions and does not affect the main conclusions of this study.

Table 3.4.2: Estimation results (including the controls for socioeconomic status)

Dependent variable: Cause-specific mortality risk	**Cancer**		**CVD**		**Other diseases**	
Covariate[a]	Parameter estimate	Standard error	Parameter estimate	Standard error	Parameter estimate	Standard error
Age	0.258	(0.031)	0.311	(0.041)	0.406	(0.051)
Birth cohort	-0.011	(0.023)	-0.105	(0.028)	-0.057	(0.028)
Woman	-1.109	(0.171)	-1.474	(0.214)	-1.610	(0.221)
Single at age 65	0.436	(0.184)	0.892	(0.211)	1.125	(0.217)
Widowed at age 65	0.516	(0.200)	0.419	(0.248)	0.821	(0.243)
Logarithm of income at age 65	-0.307	(0.146)	-0.958	(0.187)	-0.855	(0.191)
Homeowner at age 65	-0.254	(0.130)	-0.276	(0.155)	-0.555	(0.160)
Years self-employed between ages 58 and 64	0.025	(0.115)	-0.099	(0.161)	0.010	(0.155)
Years unemployed between ages 58 and 64	-0.010	(0.083)	-0.115	(0.103)	-0.109	(0.099)
Years on disability between ages 58 and 64	0.219	(0.094)	0.228	(0.099)	0.293	(0.093)
Years nonparticipating between ages 58 and 64	-0.031	(0.079)	-0.179	(0.109)	-0.161	(0.110)
Years in early retirement between ages 58 and 64	-0.009	(0.057)	-0.113	(0.071)	-0.096	(0.073)
Standard deviation random effect, σ	2.866	(0.120)				
α_1	0.832	(0.065)				
α_2	1.030	(0.093)				
α_3	1.139	(0.114)				
Log-likelihood value	-7916.5					
Number of individuals	9618					

[a] The model also includes a constant and dummy variables for the different labor market statuses at age 58 (see section 3.3).

Table 3.4.3: Cause-specific mortality by age, conditional on being alive at age 65 for a reference individual[a] (based on table 3.4.2 results)

	All causes		**Cancer**		**CVD**		**Other diseases**	
Probability of having died	%	S.e.	%	S.e.	%	S.e.	%	S.e.
by age 70	2.90	(0.63)	1.43	(0.43)	0.94	(0.43)	0.53	(0.29)
by age 75	13.25	(2.67)	5.15	(1.36)	4.17	(1.38)	3.93	(1.46)
by age 80[b]	51.04	(7.68)	16.31	(3.84)	14.51	(4.39)	20.22	(5.40)

[a] The reference individual is a male, born in 1931, married at age 65, with a median standardized household income, living in a rented accommodation at age 65, and who remained employed up to the statutory retirement age of 65.
[b] This is an out-of-sample prediction (the oldest individual in our sample is 76 years of age).

of having died by age 75, due mainly to a lower mortality risk from CVD.[19]

The probability of having died by age 75 is almost twice as high among individuals in the first quartile of the income distribution than among individuals in the fourth quartile.[20] In line with previous studies, this difference is strongest for the mortality risk from CVD but insignificant for that from cancer. Furthermore, homeowners have a 3.6 percentage point lower probability of having died by age 75, primarily due to a lower mortality risk from other diseases.

To assess the impact of pathways to statutory retirement on cause-specific mortality risk, we examine in the second last panel at the bottom of table 3.4.4 the difference in the probability of having died by age 75 between individuals who are and who are not employed in the three years preceding statutory retirement (i.e., at ages 62–64).[21] We find, for example, that individuals in early retirement during the three years preceding statutory retirement have no increased probability of having died by age 75, no matter the cause of death. The only pathway that is significantly associated with the probability of having died by age 75 is being on disability during the three years preceding statutory retirement. Compared to those who remain employed, these individuals have an almost twofold increase in the probability of having died by age 75 from any of the three causes of death. Although the level of significance of the estimated increased CVD mortality risk is somewhat lower, we still reject a one-sided hypothesis of no positive association.

[19]This interpretation only holds if we assume no time effects.
[20]Calculation: $13.25\% + 4.66\%$ versus $13.25\% - 3.86\%$.
[21]Three years is an arbitrary choice; we could equally choose, for instance, two or four years.

Finally, the last panel (at the bottom of table 3.4.4) shows that without controlling for socioeconomic status and time-invariant unobserved individual characteristics, the main findings are largely unchanged (and also considering the standard errors). There is only one exception. In the last panel the pathway nonparticipation is negatively associated with mortality risk, while there is no significant association in the second last panel. Together with the negative association between socioeconomic status and mortality risk as shown in the upper part of table 3.4.4, this could be explained by individuals with a higher observed socioeconomic status being more likely to exit the labor force to nonparticipation.

Table 3.4.4: Cause-specific mortality probabilities by socioeconomic status and pathway to retirement

	Cause-specific probability of having died by age 75 (conditional on being alive at age 65)							
	All causes		**Cancer**		**CVD[a]**		**Other diseases**	
	%	S.e.	%	S.e.	%	S.e.	%	S.e.
Reference indiv.	13.25	(2.67)	5.15	(1.36)	4.17	(1.38)	3.93	(1.46)
Differences from the reference individual								
	% point	S.e.	% point	S.e.	% point	S.e.	% point	S.e.
Gender								
Man[b]	0.00		0.00		0.00		0.00	
Woman	-9.24	(1.92)	-3.23	(0.86)	-2.98	(0.97)	-3.03	(1.11)
Marital status								
Married[b]	0.00		0.00		0.00		0.00	
Single	13.95	(3.38)	2.25	(1.25)	4.68	(1.99)	7.02	(2.19)
Widowed	9.03	(3.22)	2.65	(1.31)	1.79	(1.43)	4.59	(2.19)
Birth cohort								
1931[b]	0.00		0.00		0.00		0.00	
1936	-2.81	(1.27)	-0.21	(0.55)	-1.56	(0.68)	-1.04	(0.62)
1941	-4.63	(2.09)	-0.38	(1.03)	-2.49	(1.02)	-1.76	(0.99)
Household income								
1st quartile	4.66	(1.23)	0.65	(0.37)	2.14	(0.74)	1.87	(0.74)
Median[b]	0.00		0.00		0.00		0.00	
4th quartile	-3.86	(1.04)	-0.72	(0.38)	-1.63	(0.55)	-1.51	(0.62)
Accommodation								
Renter[b]	0.00		0.00		0.00		0.00	
Homeowner	-3.55	(1.26)	-1.00	(0.61)	-0.89	(0.57)	-1.66	(0.67)
Labor market status at ages 62-64, conditional on being employed until age 62								
Employed[b]	0.00		0.00		0.00		0.00	
Self-employed	0.03	(3.71)	0.66	(1.89)	-0.97	(1.60)	0.34	(2.12)
Unemployed	-2.07	(2.32)	0.05	(1.17)	-0.97	(1.00)	-1.15	(1.04)
On disability	12.67	(3.77)	4.53	(2.16)	3.21	(1.82)	4.93	(1.55)
Nonparticipating	-3.26	(2.41)	-0.32	(1.10)	-1.49	(1.01)	-1.45	(1.07)
Early retired	-2.14	(1.82)	-0.01	(0.81)	-1.18	(0.82)	-0.95	(0.90)
Labor market status at ages 62-64, conditional on being employed until age 62 (based on the model excluding random effects and socioeconomic status variables, table 3.4.1)								
Employed[c]	0.00		0.00		0.00		0.00	
Self-employed	0.48	(4.75)	1.33	(2.60)	-1.96	(4.68)	1.13	(5.10)
Unemployed	0.18	(2.68)	1.40	(2.08)	-0.95	(2.96)	-0.24	(2.63)
On disability	17.10	(2.54)	5.09	(2.71)	4.93	(3.59)	7.68	(3.60)
Nonparticipating	-5.29	(2.64)	0.76	(1.66)	-3.50	(2.50)	-2.65	(2.52)
Early retired	-1.70	(1.80)	1.48	(1.19)	-2.31	(2.19)	-0.87	(1.96)

[a] CVD = cardiovascular diseases

[b] The reference individual is a male, born in 1931, married at age 65, with a median standardized household income, living in a rented accommodation at age 65, and who remained employed up to the statutory retirement age of 65.

[c] The reference individual is a male, born in 1931, who remained employed up to the statutory retirement age of 65.

3.5 Summary and conclusions

In this study we investigate whether older workers who leave employment have health conditions that may prevent them from remaining employed. For this purpose we analyze to what extent pathways to statutory retirement other than employment are associated with adverse health conditions, as measured by increased cause-specific mortality risk during retirement (i.e., from the statutory retirement age of 65 onward). In addition, we provide empirical evidence for the Netherlands on the associations between socioeconomic status variables, such as marital status and income, and cause-specific mortality risk. Hereby we distinguish between cancer, cardiovascular and other diseases mortality risks.

The analysis involves the estimation of a dependent competing mortality risks model using administrative data from the 1996–2007 waves of the Income Panel Study of the Netherlands supplemented with the Causes of Death registry. Hereby we condition on both observed and unobserved individual characteristics. A methodological contribution is the use of multiple causes of death statistics in the context of a competing risk model to allow the impact of the unobserved individual characteristic to differ across the cause-specific mortality risks. We find a significant role of unobserved individual specific characteristics for all three cause-specific mortality risks and conclude that there is dependence between the three competing risks (after controlling for observed characteristics).

Our primary empirical findings can be summarized as follows. Compared to older workers who remain employed up to the statutory retirement age of 65, we find no increased mortality risk among workers who leave employment between the ages of 58 and 65 to early retirement and unemployment. Mortality risk is significantly higher among workers who leave employment between the ages of 58 and 65 and start drawing disability insurance benefits. These latter individuals have, by definition, health conditions that limit their work capacity.[22] These primary findings are not affected by allowing for random effects or the inclusion of socioeconomic status variables that are, in turn, shown to be strongly associated with mortality risk.

As discussed in the introduction, the results from previous studies support

[22]For a critical discussion on the functioning of the Dutch disability insurance scheme we refer to De Vos et al. (2010).

our assumptions of a positive association between health conditions that limit a person's work capacity and later life mortality risk, and that leaving employment to early retirement or unemployment has no health-preserving effect. Under these assumptions, our primary empirical findings imply that workers who leave employment before the statutory retirement age of 65 to early retirement or unemployment have, on average, no worse health conditions that may limit their work capacity than workers who remain employed up to age 65.

The recent reforms in the Netherlands of the early retirement and unemployment insurance schemes are aimed at keeping older workers employed (see, e.g., De Vos et al. 2010). The success of such policies depends, among other things, on the health conditions of these workers. If, for instance, workers for health reasons have chosen to retire early, the effectiveness of a policy that limits early retirement opportunities will be lower than if workers who have chosen early retirement have no health conditions that prevented them from remaining employed. Our findings imply that the effectiveness of reforms of the early retirement and unemployment insurance schemes for older workers, may, on average, not be adversely affected by the health conditions of older workers, other than those who qualify for disability insurance benefits.

4 The Effects of Cooperation; A Structural Model of Siblings' Caregiving Interactions

This chapter is based on Knoef and Kooreman (2011).

4.1 Introduction

When parents age, their adult children usually face deteriorating parental health and an increased need for care. For the children, the question arises of how to balance the goal of appropriately caring for parents with other goals in life, such as work and their own family. Governments, on the other hand, face the challenge of how to reconcile the conflicting goals of encouraging the provision of care for the elderly by families, and encouraging (female) participation in the labor market.

A prerequisite for designing effective policies in this area is to understand the complex decision making process at the level of individual families. The outcome of the decision making process depends on a large number of factors, including the labor market potential of each adult child in the family, their own family situation, the availability of formal care, the distances between the parental home and each child's home and the health status of the parents. An additional important factor that has received only scant attention in the literature is the nature of the interactions between siblings, in particular whether these can be characterized as cooperative or non-cooperative.

The purpose of this chapter is to analyze this complex process by developing a structural model in which adult children allocate their time to work, leisure, and care simultaneously. Our first contribution to the literature is that we estimate a structural model for children without siblings (only children), to learn about the preferences of adult children for informal care, without having to make assumptions about the nature of interactions between siblings. Thus our maintained assumption is that differences in behavior between children with and without siblings are due to dissimilar constraints only. In the model, preferences are characterized by a utility function defined over consumption, leisure, and the amount of care that parents receive from their children. Children face a time constraint and a budget constraint, which depend on the (potential) wage in the labor market, and the time and monetary costs of traveling to the parental home. As far as we know, this is the first study that extracts preferences with regard to informal care using only children, such that the results are not affected by interactions between siblings. Only Kotlikoff and Morris (1990) explicitly consider only children, but they analyze the living arrangements of an only child and a single parent. This study, instead, focuses on care arrangements, taking living arrangements as given.[1]

Our second contribution to the literature is a first attempt to assess the nature of the interactions between siblings and investigate the potential welfare gains of cooperation between siblings. In the literature, siblings are often ignored in the decision making process, or included only as an explanatory variable. However, as noted, among others, by Checkovich and Stern (2002), caregiving decisions among siblings are not independent and allowing for simultaneous decision making among siblings improves our understanding of caregiving decisions. The next question that arises is how these family decisions take place. Some studies that consider siblings assume that decisions are made non-cooperatively (Byrne et al., 2009, Callegaro and Pasini, 2008, Fontaine et al., 2009, Hiedemann and Stern, 1999), while others assume a two-stage decision process in which siblings (1) decide whether to participate in caregiving or not, and (2) those who participate in caregiving make a cooperative care decision (Engers and Stern, 2002). This study computes cooperative as well as non-cooperative equilibria between siblings using the estimated preference parameters from the structural model, and compares these equilibria to the observed outcomes found in the data. To do this, we have to make some assumptions. First, as mentioned before, siblings are assumed to

[1]There are some studies that model both care and living arrangements, e.g. Hoerger et al. (1996) and Pezzin and Schone (1999).

have the same preferences as only children with regard to leisure, consumption, and the amount of informal care received by the parent. Secondly, we assume that informal care provided by oneself or by a sibling are perfect substitutes. Finally, we assume that siblings have their own time and budget constraints and that there are no financial transfers between siblings.

We bring the model to the data using the first two waves of the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE includes information on the distances between the parental and adult children's homes, labor market participation, the household situation of adult children and their parents, and the amount of time spent on caring for parents. Sources of identification of the econometric model include shocks in the health condition of parents between the two SHARE waves, and variation in characteristics and outcomes between waves and between adult children. SHARE does not contain wage and income data of the adult children. Therefore, we use the European Union Statistics on Income and Living Conditions (EU-SILC) as additional data to impute wage rates and other household income for the adult children.

The results show that 71% of the siblings have a higher probability to behave non-cooperatively than cooperatively. If it is possible to push these families into their cooperative equilibrium, the amount of informal care can be increased, but this is at the expense of labor supply.

The chapter proceeds as follows. Section 4.2 discusses relevant literature on informal care giving. In section 4.3 we specify the structural model and explain the estimation strategy. Section 4.4 discusses the data, after which section 4.5 presents the estimation results. Section 4.6 considers the nature of the interactions between siblings (cooperative and non-cooperative equilibria) and investigates the potential welfare gains of cooperation. Section 4.7 concludes.

Literature review 4.2

In the economic, demographic, sociological, and psychological literature on the elderly, considerable attention has been paid to the degree to which children support their (elderly) parents. Support itself is usually distinguished into instrumental support on the one hand, and social and emotional support on the other (Hogan and Eggebeen, 1995, Silverstein and Bengtson, 1997). This study focuses on instrumental support, which includes practical help to parents (e.g., running errands, doing household work), help with personal care (e.g., washing, bathing, care when sick) and help with paperwork. Research shows that children often provide practical help to their parents. Even in later life, however, parents in Europe more often help children than children help parents (Kohli, 1999). Hence, there is little reversal of the flow of practical support exchange as parents age.

Another category of instrumental support is financial support. Financial support to parents is rarely given by children in western societies, except among immigrants. Bonsang (2007) found that only 2.6% of adult children in European countries provide financial assistance to their parents. In non-western societies, it is more common and often more obligatory that adult children financially support their parents (Frankenberg et al., 2002, Lee et al., 1994). Financial support from parents to children is more common. However, these financial transfers are mainly to children following further education or less well off children, such as those who are unemployed. As these motivations are not directly related to informal care giving, this study does not take financial transfers explicitly into account.

In the empirical economic literature we find reduced form models and structural models investigating (1) the extent to which informal care and formal care are complements or substitutes, (2) the factors that determine the provision of informal care, and (3) the dependence between informal care giving and labor supply.

If informal and formal care are substitutes, informal care can reduce home health care use and delay nursing home entry. Only then, governmental long term care expenditures can be reduced and labor shortages in the (long term) health care sector can be reduced, by increasing informal care. Bolin et al. (2008a) and Bonsang (2009) investigated this issue in European countries and found that informal care is a substitute for long term care, at least as long as the needs of the elderly are low and require unskilled types of care. For the U.S. Van Houtven and Norton (2004) also conclude that informal care and formal care are substitutes. On the other hand, the introduction of free formal personal care in Scotland in 2002 does not seem to have reduced informal care (Bell et al., 2006).

The models in the literature focus on a large number of potential determinants. Theoretically, these determinants can be distinguished into demand and supply variables. Demand variables are characteristics of parents which indicate the degree to which parents 'need' support from their children, such as a parent's health status, and whether the parent is living with a partner (Grundy, 2005, Klein Ikkink et al., 1999, Silverstein, 1995, Spitze and Logan, 1989). Living with a partner is related to less need for support by children, because the partner is the prime source of giving support to an elderly person (Dykstra, 1993).

Supply variables have to do with the child's costs and benefits of giving support. Research shows that there is variation among societies in the degree to which children respond to the need of their parents, with children in individualistic countries like Sweden and the Netherlands being less responsive (Kalmijn and Saraceno, 2008). We will therefore include country specific dummy variables to allow the preferences for informal care to differ across countries.

An important supply variable is time costs. Giving support and paying a visit are time intensive, especially if support also requires traveling, which is usually the case. There are also financial costs involved, but there is little evidence that the child's income situation affects contact or support (Klein Ikkink et al., 1999, Waite and Harrison, 1992). There are social status gradients in contact and support, but these have more to do with education and less with financial aspects of social status (Kalmijn and Dykstra, 2006).

The time budget of an adult child depends on whether the child has children. Several authors have hypothesized that caring for one's own children competes with the support children give to their elderly parents. This phenomenon has been referred to as the 'sandwich generation'. There is indeed some evidence that the support daughters give to parents is negatively affected by having children (Klein Ikkink et al., 1999), but there is also evidence for a null effect (Eggebeen and Hogan, 1990). A complication is that having one's own children may also increase contact levels with the parent due to the grandparenting role (Kalmijn and Dykstra, 2006). This may be a reason why there are no consistent effects of having children on support.

Employment also affects children's time budget, and the opportunity costs of labor may influence the informal care decision. Several studies have investigated the relation between employment and informal care using different datasets and methods to correct for the potential endogeneity bias (caregivers may have different (unobserved) characteristics than non-caregivers, which influence both informal care and labor market decisions). The results are mixed. Wolf and Soldo (1994) find no evidence of reduced propensities to be employed, or of reduced conditional hours of work, due to the provision of informal care. Others find that informal care reduces employment significantly among European men and women (Bolin et al., 2008b), and among U.S. women (Ettner, 1996). Ettner (1995) and Heitmueller and Michaud (2006) find that caregiving for coresidential parents reduces employment. As in Pezzin and Schone (1997, 1999), Byrne et al. (2009), and Callegaro and Pasini (2008) we will model the labor force decision and informal care decision jointly in a structural model. The results are important for understanding the conflict between women's increasing economic role in society on the one hand, and the increasing need for informal support to the elderly on the other (Kohli, 1999).

A final determinant of informal care has to do with family size and family interactions. The number of siblings in a family may have different effects. First, parents will need less help from each individual child when they have more children. In addition, children may shirk their responsibilities if there are many siblings who can do the work, such that the amount of informal care given by one sibling may depend negatively on the care provided by another. On the other hand, in the case of a strategic bequest motive (described by Bernheim et al., 1985), the amount of care given by a sibling depends positively on the care given by the other siblings. However, more recent studies do not support the bequest motive (Callegaro and Pasini, 2008, Perozek, 1998, Sloan et al., 1997). It has been found that siblings are each other's substitutes. The more siblings a child has, the less often the child visits the parent and the less often he or she gives support to the parent (Kalmijn, 2007, Kalmijn and Saraceno, 2008, Spitze and Logan, 1991). In addition to the number of siblings, the nature of the interactions between siblings plays a role in informal care decisions. In the literature we do not find evidence regarding whether siblings behave cooperatively or non-cooperatively. This study tries to identify the behavior of siblings using the preference parameters of only children which are obtained in a structural model.

Structural model 4.3

This section describes the structural model we use to estimate the amount of time only children spend on providing informal care to their parents, taking into account the key supply and demand factors discussed in the previous section. Section 4.3.1 deals with the specification of the model and describes the estimation strategy. Section 4.3.2 explains how we impute wage rates and other household income in the model, because SHARE contains no information about the wage rates and other household income of the adult children. We use a wage equation to impute wage rates and an income equation to impute remaining household income for the adult children in SHARE.

Model specification 4.3.1

We specify a structural model to explain the amount of time an adult child spends on paid work, care for parents, and other activities. In this study all activities other than paid work and care for parents are called leisure. As in Van Soest (1995), we formulate the model as a discrete choice problem. In this discrete choice problem adult children can choose between different combinations of labor, informal care, and leisure, which also lead to different levels of consumption. With regard to labor we distinguish full-time employment, part-time employment, and no employment.[2] In the model, full-time employment is set to 36 hours of labor per week and part-time employment to 18 hours of labor per week. Concerning informal care, we consider the choice to give no substantial amount of informal care, giving between 1 and 4 hours a week (50% of the informal care givers), between 4 and 8 hours a week (20%) and giving more than 8 hours of informal care a week (30% of the informal care givers). Where no substantial amount of informal care is given, the hours of informal care in the model is set to zero.[3] For the second informal care category (1–4 hours) we set the number of hours of informal care in the model to be 2 (the average) and the number of visits to one per week, for the second category (4–8 hours) the number of hours is six (the average) and visits are on

[2]These are the three categories available in the data.

[3]In the data there are 134 observations giving between 0 and 1 hour of informal care per week. Most of them give less than 0.25 hours of informal care per week. These people fall into the category ‘no substantial informal care’.

a daily basis.[4] In the last category (> 8 hours per week) we set the number of hours of informal care to be 18[5] and we assume that the parents are visited on a daily basis, which is also the median number of visits in this category. In total we thus have a choice set of 12 alternatives (3 labor market categories × 4 informal care categories).

The child derives utility from leisure (t_l), consumption (c), and the amount of informal care his parents receive (t_s). We use the following quadratic utility function

$$U(t) = t'At + t'b, \tag{4.1}$$

where $t = (t_l, c, t_s)'$, A is a symmetric 3×3 matrix with entries $\alpha_{ij}(i, j = 1, 2, 3)$ and $b = (b_l, b_c, b_s)'$. For the model to be economically rational, the marginal utility of consumption must be positive; see e.g. Van Soest and Stancanelli (2010). We will check whether this condition is satisfied in its estimated version. The marginal utility of informal care may be negative.[6] We maximize the utility function subject to a time and budget constraint. The time and budget constraints are specified as

$$\begin{aligned} & t_l + t_h + t_s + (\tau d)K = T \\ & c + K p_d d = w t_h + \mu \end{aligned} \tag{4.2}$$

where

t_h = labor time (hours),

K = number of visits (per week),

d = distance to parent (return trip, km),

τ = travel time per kilometer (hours),

T = total time (# hours in one week),

[4] The median number of visits in the 4–8 category is also seven per week.

[5] This is the median number of hours of informal care in the '> 8' category. The average number of hours of informal care in this category is 29, but this is due to some individuals giving a very high number of hours of informal care.

[6] Estimates of Byrne et al. (2009) show that adult children care about their parents' health quality, suggesting that altruism may play an important role in the provision of informal care. However, they also show that informal care provision tends to be burdensome, which may explain why few family members provide care for elderly individuals.

$$p_d = \text{travel costs (per kilometer)},$$
$$w = \text{wage (per hour)},$$
$$\mu = \text{remaining household income.}$$

The time endowment T is 168 hours per week. Remaining household income (μ) includes all income that is not earned by the adult child under consideration. It includes capital income, social transfers, and labor income of the partner (if present). We abstract from the fact that labor market choices of the adult children under consideration and their partners may be determined simultaneously. Furthermore, we assume wage rates[7] and the geographical distance between adult children and their parents to be exogenous.[8]

To take into account preference variation across adult children, the vectors in b are functions of observed and unobserved characteristics of the adult children and their parents

$$\begin{aligned} b_l &= X_l\beta_l + u_l \\ b_c &= X_c\beta_c + u_c \\ b_s &= X_s\beta_s + u_s. \end{aligned} \tag{4.3}$$

X_l and X_c contain characteristics which are likely to influence the amount of leisure time and consumption the adult child prefers, such as the age, gender, education, number of children, and marital status of the adult child. X_s includes variables influencing the preference for giving informal care to parents, namely the health position of the parents, whether both parents are alive and the gender of the parent when the parent is single, the (average) age of the parents, the gender of the child, country specific dummy variables, and the number of children of the adult child. Also education is included in the matrix X_s, because higher educated children may have different value orientations (Kalmijn, 2006). Random preferences due to unobserved characteristics are incorporated through the terms u_l, u_c, and u_s. They capture time invariant unobserved heterogeneity. For example, u_s may capture the three motives

[7]Bolin et al. (2008b) found no statistically significant wage-rate effects of informal care provision in Europe.

[8]Charles and Sevak (2005) tested whether children's location endogenously responds to parent's health but found no evidence of this.

that are, in addition to observed characteristics, important in explaining social support: reciprocity, altruism, and norms of responsibility.[9] We assume $u = (u_l, u_c, u_s)$ to be distributed jointly normal with mean zero and covariance Σ_u

$$\begin{bmatrix} u_l \\ u_c \\ u_s \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_l^2 & \sigma_{l,c} & \sigma_{l,s} \\ \sigma_{l,c} & \sigma_c^2 & \sigma_{c,s} \\ \sigma_{l,s} & \sigma_{c,s} & \sigma_s^2 \end{bmatrix} \right). \tag{4.4}$$

In addition, we introduce random disturbances to the utilities of the twelve choice opportunities in the same way as in the multinomial logit model

$$\begin{aligned} &U_j = U(t_l, c, t_s) + \varepsilon_j \quad j = 1, ..., 12 \\ &\varepsilon_j \sim EV(I) \quad j = 1, ..., 12 \quad \varepsilon_1, ..., \varepsilon_{12} \text{ independent} \end{aligned} \tag{4.5}$$

leading to the familiar logit choice probabilities

$$\begin{aligned} &P\left(U_j > U_k \text{ for all } k \neq j \mid X, d, w, \mu, u\right) \\ &\qquad = \frac{\exp(U(t_j))}{\sum_{k=1}^{12} \exp(U(t_k))}. \end{aligned} \tag{4.6}$$

Substituting the utility function (4.1) and the time and budget constraint (4.2), equation (4.6) becomes

$$\begin{aligned} &P\left(U_j > U_k \text{ for all } k \neq j \mid X, d, w, \mu, u\right) \\ &\qquad = \frac{\exp\left(t_j' A t_j + t_j' b\right)}{\sum_{k=1}^{12} \exp\left(t_k' A t_k + t_k' b\right)}, \end{aligned} \tag{4.7}$$

where $t_j = \left(t_{lj}, c_j, t_{sj}\right)$ and t_{lj} and c_j are defined by

$$\begin{aligned} t_{lj} &= T - t_{hj} - t_{sj} - (\tau d) K_j \\ c_j &= w t_{hj} + \mu - K_j p_d d. \end{aligned} \tag{4.8}$$

[9]These three motives are investigated in the sociological literature (e.g. Kohli and Künemund, 2003, and Kalmijn, 2010). Kalmijn (2010) found that altruism is relatively important for parents to support their children, however, for adult children, reciprocity and norms of responsibility appear to be relatively more important.

Equation (4.7) presents the probability that a certain combination of (t_l, c, t_s) is chosen, given observed and unobserved characteristics. The disturbances ε_j can be interpreted as optimization errors: adult children choose a combination of (t_l, c, t_s) that is close to optimal, rather than always fully optimal. This may be due to errors in the perception of the utilities of the set of alternatives. In contrast, the random effects (u_l, u_c, u_s) are known by the adult child (but unobserved to the researcher). We estimate the model parameters using maximum likelihood. The likelihood contribution of an individual i who chooses alternative j is

$$\begin{aligned} L_i(\alpha, \beta, \Sigma_u \mid X, d, w, \mu) \\ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(U_j > U_k \text{ for all } k \neq j \mid X, d, w, \mu, u) p(u) du, \end{aligned} \tag{4.9}$$

where $p(u)$ is the density of vector u. The three dimensional integral can be approximated using simulations (simulated maximum likelihood). Using R simulations, the likelihood contribution of equation (4.9) becomes

$$L_{iR}(\alpha, \beta, \Sigma_u \mid X, d, w, \mu) = \frac{1}{R} \sum_{r=1}^{R} P(U_j > U_k \text{ for all } k \neq j \mid X, d, w, \mu, u^r), \tag{4.10}$$

where the draws $u^r, r = 1 \ldots R$ are from a trivariate normal distribution with mean zero and variance Σ_u. Most of the adult children are observed twice (wave 1 and wave 2). The likelihood contribution of an adult child who is observed in both waves, and chooses alternative j in wave 1 and alternative h in wave 2 is

$$\begin{aligned} L_{iR}(\alpha, \beta, \Sigma_u \mid X, d, w, \mu) = \\ \frac{1}{R} \sum_{r=1}^{R} P(U_{j1} > U_{k1} \text{ for all } k \neq j \mid X_1, d_1, w_1, \mu_1, u^r) \\ \times P(U_{h2} > U_{k2} \text{ for all } k \neq h \mid X_2, d_2, w_2, \mu_2, u^r), \end{aligned} \tag{4.11}$$

so that the unobserved characteristics are the same in both waves.

A draw u^r can be obtained by taking 3 (pseudo-random) draws from a standard normal distribution (which we shall call $\theta = (\theta_l, \theta_c, \theta_s)'$) and then calculating $\left(u_l^r, u_c^r, u_s^r\right)' = L\theta$. Here, L is the Choleski factor of Σ_u (the unique lower triangular matrix such that $LL' = \Sigma_u$).[10]

Integrals can be approximated with fewer draws (R) when using Halton draws instead of pseudo-random draws. This is because Halton sequences provide more coverage of the density which has to be integrated. For more information about the derivation of Halton sequences see for example Train (2003), or Drukker and Gates (2006), who discuss the advantages of Halton sequences when using simulations to approximate integrals numerically.

4.3.2 Modeling wage rates and remaining household income

Wage rates (w) and remaining household income (μ) of the adult children in SHARE are unknown. Therefore, we use predictions from a wage equation[11] and an equation for remaining household income. Both equations are estimated using the 'European Union Statistics on Income and Living Conditions' (EU-SILC).

In EU-SILC we can only observe wages for workers. However, the working population is probably not a random subsample from the population as people with comparatively high wages (conditional on, for example, their education level) are more likely to work. There may be unobservables that influence the decision to participate, as well as the wage rate. A commonly used method to deal with this sample selection is the method presented by Heckman (1979). Heckman takes selection bias into account by adding an equation which models the participation decision, and allowing for nonzero correlation between the wage and the participation equation. We estimate the following Heckman model, for each country separately

$$\ln(w_i^*) = X_{wi}\beta_w + v_{wi} \tag{4.12a}$$
$$p_i^* = X_{pi}\beta_p + v_{pi} \tag{4.12b}$$

[10] u is normally distributed because the sum of normals is normal. Furthermore, the covariance of u is Σ_u because $\text{Var}(u) = E(uu') = E(L\theta(\theta L)') = LE(\theta\theta')L' = L\text{Var}(\theta)L' = LIL' = LL' = \Sigma_u$ (Train, 2003).

[11] We assume wage rates to be independent of the provision of informal care. This is consistent with the results of Bolin et al. (2008b), who did not find any statistically significant wage-rate effects of informal care provision.

$$w_i = w_i^* \quad \text{if } p_i^* > 0 \tag{4.12c}$$

$$w_i = 0 \quad \text{if } p_i^* \leq 0 \tag{4.12d}$$

where (4.12a) is the wage equation and (4.12b) is the (probit type) participation equation. X_{wi} and X_{pi} contain personal characteristics such as age, gender, and education level. Generally an exclusion restriction is required to generate credible estimates from the Heckman selection model. Therefore, we include dummy variables for having children in the participation equation, but exclude these from the wage equation. We assume that v_p and v_w are bivariate normal distributed

$$\begin{bmatrix} v_p \\ v_w \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{wp} \\ \sigma_{wp} & \sigma_w^2 \end{bmatrix}\right) \tag{4.13}$$

and we estimate the parameters using FIML. As for a probit model, the normalization $\sigma_p^2 = 1$ is used since only the sign of p_i^* is observed. For remaining household income (μ), we also estimate an equation using a standard OLS regression, for each country and for men and women separately

$$\ln(\mu_i) = X_{\mu i}\beta_\mu + v_{\mu i}, \tag{4.14}$$

where $X_{\mu i}$ contains personal characteristics such as age, marital status, and education level.

In the structural model, introduced in section 4.3.1, we take into account that wage rates and remaining household income are predicted with error. Using the estimated variances of the errors in the wage equations and the remaining household income equations (σ_w^2 and σ_μ^2), we integrate the prediction errors out. Van Soest (1995) also uses estimated standard deviations of the errors in the wage equation to account for prediction errors.

When we take into account prediction errors, the likelihood contribution in equation (4.9) of an individual who chooses alternative j becomes

$$\begin{aligned} &L\left(\alpha, \beta, \Sigma_u \mid X, d, \beta_w, \sigma_w, \beta_\mu, \sigma_\mu\right) \qquad (4.15)\\ &= \int\!\!\int\!\!\int\!\!\int\!\!\int_{-\infty}^{+\infty} P\left(U_j > U_k \text{ for all } k \neq j \mid X, d, w, \mu, u\right) p(u)p(w)p(\mu)du dw d\mu. \end{aligned}$$

So that equation (4.10) becomes

$$L_{iR}\left(\alpha,\beta,\Sigma_u \mid X,d,\beta_w,\sigma_w,\beta_\mu,\sigma_\mu\right) = \frac{1}{R}\sum_{r=1}^{R} P\left(U_j > U_k \text{ for all } k \neq j \mid X,d,w^r,\mu^r,u^r\right), \tag{4.16}$$

where

$$w^r = \exp\left(X_w'\beta_w + v_w^r\right) \tag{4.17}$$

and v^r is a draw from the normal distribution with variance σ_w^2. In the same way

$$\mu^r = \exp\left(X_\mu'\beta_\mu + v_\mu^r\right), \tag{4.18}$$

where v_μ^r is a draw from the normal distribution with variance σ_μ^2.

For most countries the estimates of σ_{wp} in the EU-SILC data are not significant, which indicates that selection with regard to unobservables is not very important. We therefore do not take into account correlations between v_w, v_μ and the unobserved characteristics (u_l, u_c, u_s).

4.4 Data

This section describes the data we use to estimate the parameters of the model. Section 4.4.1 describes the Survey of Health, Ageing and Retirement in Europe (SHARE) and section 4.4.2 the 'European Union Statistics on Income and Living Conditions' (EU-SILC).

4.4.1 SHARE

SHARE is a multidisciplinary database of microdata on health, socio-economic status and social and family networks of individuals aged 50 and older in Europe. Data were collected in 2004/2005 (wave 1) and 2006/2007 (wave 2) by face-to-face computer-aided personal interviews (CAPI), plus a selfcompletion drop-off part with questions that require more privacy. This study uses 12 countries that have contributed data to SHARE. They represent various regions in Europe, ranging from Scandinavia (Denmark and Sweden) through Central Europe (Austria, France, Germany, Belgium, and the Netherlands) to the Mediterranean (Spain, Italy and Greece). In the second wave two 'new' EU member states have contributed data to SHARE (Czech Republic and Poland). Other countries available in SHARE that we do not use in this study are Israel and Ireland. We do not use these countries because they are not represented in the EU-SILC data, which we describe in the next section.

Several papers use SHARE to study informal care giving. Most of these studies use the respondents as providers of informal care (e.g. Bonsang, 2007, 2009, and Bolin et al., 2008a,b). This study considers the respondents in their role as (the potential) receiver of informal care. Crespo and Mira (2010) call this the 'parents-sample' as the respondents are the elderly parents. The reason for using the 'parents-sample' is that we need information on all siblings within a family. The respondents (in our case 'the parents') give information about all their children that are still alive (sex, year of birth, geographical distance between the children and their parents, education, marital status, number of children, the employment status of the children, and the amount of informal care they receive from their children). If we were to consider the respondents as the providers of informal care, there would be no information on the amount of care the siblings of the respondents give to their parents. The health situation of the parents provides a measure for the amount of care parents need. SHARE provides a lot of health related variables, such as self-reported health, limitations in activities of daily living (ADL and IADL), mental health, diagnosed chronic conditions, whether people are suffering from several symptoms and limitations in functioning (e.g. measures by grip strength and walking speed). In this study we use self-reported health which has the lowest number of missing data. The parents are asked to rate their health on a five-point scale, ranging from very good to very poor (wave 1) or from excellent to poor (wave 2).

We select all respondents with one or two adult children. Furthermore, our interest is in children who are 40 years or older, as these children are most likely to be involved in personal care for their elderly parents. Following McGarry (1999), Bonsang (2007), and Norton and Van Houtven (2006) we omit households where children are living in the same household as the respondent, because there is no detailed information on informal care giving within households. For the same reason we exclude respondents where grand-children, siblings, and non-relatives are living in the same household as the respondent. Families with one or two self-employed adult children are excluded, because we have no information about the number of hours that self-employed people work. Also families where one or both children have the daily activity given as 'sick' are excluded, as they may not be able to give informal care. After excluding respondents for whom key information is missing, we end up with 2253 respondents with one adult child and 2891 respondents with two adult children.

Table 4.4.1 shows the amount of informal care and the number of adult children per country. Informal care includes practical household help (e.g. household chores, shopping and home repairs), personal care (e.g. dressing, bathing, eating) and help with paperwork. Adults report whether their children help them on an almost daily basis, weekly, monthly or less often. Furthermore, they were asked to give an estimate of the number of hours of informal care received on a typical day, week, month or year. We transform these answers to a variable measuring the average amount of informal care that adults receive from their children per week. We define people as involved in informal care when they give one hour or more of informal care per week.

In Germany, Greece, the Czech Republic and Poland, many people are involved in informal care giving (more than 15% of the only children and siblings). Conditional on being involved in informal care, children in Mediterranean countries give relatively many hours of informal care, whereas the children in Denmark, the Netherlands, and Sweden give a relatively small number of hours of informal care. When we compare only children and siblings, we find that in general only children are more often involved in informal care giving than siblings and that they also provide more hours of informal care. This suggests that the hours of care provided by a sibling are a substitute for someone's own informal care.

Table 4.4.2 presents information about informal care giving and the geographical distances between children and their parents. The higher the distance between children and their parents, the higher the traveling time and costs, and the lower the fraction of people involved in informal care. It appears that

Table 4.4.1: Informal care per country[a]

Country	only child	% informal care	# hrs	1 sibling	% informal care	# hrs
Austria	218	14.7	15.9	438	12.3	6.9
Germany	294	19.0	17.3	572	15.0	6.3
Sweden	217	10.6	7.1	674	7.6	5.9
Netherlands	115	7.8	3.0	442	4.3	4.8
Spain	99	13.1	17.2	308	9.7	19.5
Italy	167	12.0	18.4	338	8.6	12.8
France	263	14.1	10.0	508	9.6	6.2
Denmark	134	11.2	4.5	512	6.3	6.8
Greece	213	19.7	17.1	804	19.5	12.5
Belgium	318	20.1	5.8	528	8.1	10.8
Czech Republic	165	24.2	11.8	450	29.6	10.7
Poland	50	16.0	16.5	208	16.8	5.1
Total	2253	15.9	12.2	5782	12.4	9.5

[a] Percentage of children involved in informal care and the number of hours of informal care, conditional on giving any informal care, per country.

the distribution of only children and siblings among the categories is about the same (so that only children do not in general live closer or further away from their parents than siblings).

As expected, the provision of informal care is higher for children with parents in bad health than for children with parents in good health (table 4.4.3). In the analysis we distinguish single parents and parents living with a partner,

Table 4.4.2: Distance and informal care[a]

Distance	only child	% inf. care	# hrs	1 sibling	% inf. care	# hrs
same building	9.8	29.0	15.7	7.1	31.0	12.4
≤ 1 kilometer	17.2	20.7	11.0	15.3	19.1	12.3
1–5 kilometers	18.8	19.9	8.1	21.0	15.9	8.3
5–25 kilometers	25.6	15.9	13.1	23.2	10.5	6.3
25–100 kilometers	12.6	9.9	13.3	15.3	6.8	5.7
100–500 kilometers	10.0	4.4	24.3	11.1	3.6	6.5
≥ 500 kilometers	3.0	0.0	-	3.3	1.1	86.0
≥ 500 kilometers and another country	3.0	1.5	1.9	3.7	1.4	1.4
Total	100	15.9	12.2	100	12.4	9.5

[a] Percentage of children involved in informal care and the number of hours of informal care, conditional on giving any informal care, per distance category.

Table 4.4.3: Health and informal care[a]

Health	only child	% inf. care	# hrs	1 sibling	% inf. care	# hrs
Father, good / very good	8.7	6.1	4.5	7.7	6.3	13.5
Father, fair	4.9	19.8	9.6	5.9	12.4	5.0
Father, poor	2.4	34.0	15.9	2.3	23.5	8.4
Mother, good / very good	21.3	16.4	7.6	22.8	12.2	7.1
Mother, fair	17.8	24.2	9.8	15.0	19.5	10.1
Mother, poor	8.5	33.3	22.0	7.1	26.5	12.1
Both poor, or poor and fair	5.0	20.5	23.3	5.3	21.8	14.6
Both fair, or fair and good	15.6	6.5	7.2	17.5	5.9	7.9
Both good / very good	12.4	2.5	7.1	13.0	3.1	3.8
Father poor, mother good	1.7	12.8	2.9	1.7	10.4	5.6
Father good, mother poor	1.6	25.0	11.8	1.8	18.3	11.1
Total	100	15.9	12.2	100	12.4	9.5

[a] Percentage of children involved in informal care and the number of hours of informal care, conditional on giving any informal care, per health status of the elderly parent. In the first three categories the adult child only has a father, in the fourth to the sixth category the adult child only has a mother, and in the last five categories the adult child has a father and a mother.

Table 4.4.4: Daily activity and informal care[a]

Daily activity	only child	% informal care	# hrs	1 sibling	% informal care	# hrs
full-time work	67.2	13.4	8.3	73.6	11.0	7.8
part-time work	8.2	15.2	7.6	8.8	11.2	5.9
unemployed	5.5	17.1	11.0	3.0	16.8	13.2
in education	0.6	7.1	14.0	0.3	0.0	-
parental leave	0.3	0.0	-	0.1	0.0	-
(early) retirement	8.1	31.1	20.4	5.4	26.4	10.9
homemaker	9.2	21.7	21.8	8.1	17.3	18.8
other	0.9	20.0	25.1	0.8	0.0	-
Total	100	15.9	12.2	100	12.4	9.5

[a] Percentage of children involved in informal care and the number of hours of informal care, conditional on giving any informal care, per daily activity of the adult child.

as parents may give informal care to each other when they are both alive. It appears that when the mother of a child is in poor health and the father is in good health there is more informal care from adult children than when the father is in poor health and the mother is in good health. The reason may be that men in the observed generations have less household management skills than women.

Table 4.4.4 shows the amount of informal care by the daily activity of the child. It is interesting to see that the amount of informal care does not differ much between children in full-time employment and those in part-time employment. Children who are (early) retired or are looking after the home are most often involved in informal care. However, note that retired persons have relatively older parents, who are more often in bad health. Finally, women are more often involved in informal care than men and often provide more hours of informal care (table 4.4.5).

Table 4.4.5: Gender[a]

Gender	only child	% informal care	# hrs	1 sibling	% informal care	# hrs
Female	53.8	17.7	14.1	51.9	14.8	9.9
Male	46.2	13.8	9.4	48.1	9.8	8.8
Total	100	15.9	12.2	100	12.4	9.5

[a] Percentage of children involved in informal care and the number of hours of informal care, conditional on giving any informal care, per gender of the adult child.

EU-SILC 4.4.2

The wage equation and the equation for remaining household income, described in section 4.3.2, are estimated using EU-SILC data. EU-SILC contains microdata on income, poverty, social exclusion and living conditions in Europe. It comprises information from surveys and registers from the EU member states. We select people up to age 76 and omit households who receive income from self-employment or who are permanently sick or disabled (just as in SHARE). Furthermore, we exclude observations which have missing data for one or more of the variables in the model. We end up with 55,100 observations, which are described in table 4.4.6.

Table 4.4.6: Descriptives EU-SILC

	AT	BE	CZ	DE	DK	ES	FR
Male (%)	49	48	47	46	49	47	48
Age (mean)	45	43	46	47	45	43	43
Primary education (%)	1	14	0	2	0	33	12
Lower secondary education (%)	24	18	19	16	29	22	13
(Upper) secondary education (%)	53	34	69	46	43	21	47
Post secondary non-tertiary education (%)	9	2	1	6	0	1	2
Tertiary education (%)	13	31	12	30	28	23	26
Man with partner (%)	35	32	32	32	38	31	34
Woman with partner (%)	34	35	32	31	37	33	32
Man with child (%)	22	24	17	21	23	24	25
Woman with child (%)	23	26	19	25	25	24	26
Net wage rate (mean)	10	11	2	10	14	8	11
Nonlabor income (mean)	27413	24240	5990	24718	28575	18717	24010
N	1488	1346	1095	6028	1422	7171	3221

	GR	IT	NL	PL	SE		Total
Male (%)	44	47	51	46	50		47
Age (mean)	43	46	45	42	43		44
Primary education (%)	28	27	9	17	9		18
Lower secondary education (%)	13	29	24	7	16		20
(Upper) secondary education (%)	36	32	37	60	42		41
Post secondary non-tertiary education (%)	5	5	3	3	5		4
Tertiary education (%)	19	7	27	13	28		18
Man with partner (%)	28	30	40	30	38		32
Woman with partner (%)	32	32	35	33	37		33
Man with child (%)	22	21	25	27	27		23
Woman with child (%)	25	22	21	31	26		25
Net wage rate (mean)	7	9	12	2	10		9
Nonlabor income (mean)	15475	22161	22036	4835	23742		18709
N	1345	14155	6007	10464	1358		55100

Estimation results 4.5

This section presents the estimation results of the wage equation, the equation for remaining household income, and the parameters of the structural model. We start with the estimation results of the wage equation and the equation for remaining household income, since these are needed as input to estimate the parameters of the structural model.

Wage equation and remaining household income 4.5.1

Wage equations are estimated for each country separately. Table 4.5.1 describes the wage equation for Sweden. The wage equations for all other countries are estimated in a similar way and are presented in Appendix 4.A. Table 4.5.1 shows that wage rates increase with age and are significantly higher for people with a high education level. σ_{wp} is not significantly different from zero, indicating that sample selection is not a significant issue. This also holds for most of the other countries. Due to measurement errors in the wage rates, the standard deviation of the errors in the wage equation may be overestimated.[12]

Table 4.5.2 shows the estimation results of remaining household income for Sweden. The equations for the other countries are estimated in a similar way and are available on request. Remaining household income increases with age. Furthermore, in Sweden remaining household income is not significantly different for different education categories. Next, we will use the wage equations and the equations for remaining household income from EU-SILC to estimate the parameters of the structural model.

Estimation results of the structural model 4.5.2

Table 4.5.3 presents the estimation results of the structural model.[13] This section first describes the parameter estimates related to the preferences for

[12]A sensitivity analysis, in which we for example multiply σ_w by 0.8 for all countries, indicates that this does not influence the structural estimation results very much.

[13]Our estimation procedure uses 25 drawings. The estimation is computer intensive. Other studies with these kind of models have used for example 5 or 10 drawings which produce qualitatively similar results (Van Soest, 1995) or 10 drawings (Van Soest and Stancanelli, 2010).

Table 4.5.1: Estimation results wage equation Sweden, sample selection model[a]

Equation 1: ln(wage rate)	Coefficient	St. error
Man	0.157	0.105
Age	0.019	0.017
$Age^2/100$	-0.010	0.020
Primary education	-0.070	0.109
Lower secondary education	-0.057	0.083
(Upper) secondary education	0.000	-
Post secondary non-tertiary education	0.051	0.089
Tertiary education	0.109	0.046
Man with partner	0.073	0.088
Woman with partner	-0.092	0.082
Intercept	1.458	0.368
Equation 2: participation decision		
Man	-0.069	0.196
Age 15–29	0.000	-
Age 30–39	1.045	0.157
Age 40–49	1.010	0.143
Age 50–59	1.147	0.166
Age $\geq$ 60	-1.193	0.155
Primary education	-0.351	0.163
Lower secondary education	-0.962	0.130
(Upper) secondary education	0.000	-
Post secondary non-tertiary education	-0.355	0.194
Tertiary education	0.090	0.117
Man with partner	0.612	0.152
Woman with partner	0.574	0.145
Man with child	-0.022	0.144
Woman with child	-0.514	0.151
Intercept	0.206	0.169
ρ	0.016	0.157
σ_w	0.615	0.014
$\sigma_{wp} = \rho\sigma_w$	0.010	0.097
N	1358	
Censored observations	422	
Uncensored observations	936	
Log likelihood	-1374.725	

[a] The reference individual is a woman with (upper) secondary education in the age category 15-29. She has no partner and no children.

Table 4.5.2: Estimation results remaining household income, Sweden[a]

ln(remaining household income)	Men		Women	
	Coefficient	St. error	Coefficient	St. error
Age	-0.097	0.023	-0.040	0.023
Age^2	0.001	0.000	0.001	0.000
Primary education	-0.115	0.227	0.051	0.232
Lower secondary education	0.326	0.186	0.185	0.189
(Upper) secondary education	0.000	-	0.000	-
Post sec. non-tertiary education	0.109	0.248	-0.156	0.300
Tertiary education	0.203	0.149	0.028	0.136
Married	0.440	0.163	0.421	0.167
Widowed	0.054	0.453	-0.442	0.341
Divorced	-0.658	0.292	-0.676	0.231
Never married	0.000	-	0.000	-
Having a child	0.724	0.138	0.845	0.141
Intercept	10.524	0.454	9.766	0.468
N	655		638	
R-squared	0.115		0.116	
Adj R-squared	0.101		0.102	
σ_μ	1.466		1.434	

[a] The reference individual is a man (left) or woman (right) who has never been married, with (upper) secondary education and no children.

informal care (t_s). With regard to informal care the results show significant decreasing returns to scale (α_{ss} is significantly negative). Furthermore, the interaction term α_{ls} is significantly positive, meaning that when the amount of informal care is already high, the utility of an extra hour of leisure increases. When parents are in bad health they need more attention and the estimates show that this increases the preference for informal care. The preference for informal care is highest when a single living father or mother has poor health, when both parents are in poor health, or when the mother has poor health and the father is in good health. On the other hand, when the father is in poor health and the mother is in good health, the preference for informal care giving is lower. Presumably, mothers are better able to give informal care to their spouses than fathers are able to give informal care to the mothers of the adult children. Several studies find that mothers receive more care than fathers (Attias-Donfut et al. 2005, Bonsang 2007, Klein Ikkink et al. 1999). Our results suggest that this depends on the health of the parent. Mothers in good health receive more informal care than fathers in good health, but fathers in bad

Table 4.5.3: Estimation results structural model[a]

		Coef.	Std. err.	p-value
α_{ll}	(t_l^2)	-0.00018	0.00019	0.358
α_{cc}	(c^2)	1.44e-07	2.38e-07	0.546
α_{ss}	(t_s^2)	-0.01226	0.00445	0.006
α_{lc}	$(t_l \times c)$	0.00002	0.00001	0.003
α_{ls}	$(t_s \times t_l)$	0.00230	0.00102	0.023
α_{cs}	$(t_s \times c)$	-1.43e-06	0.00002	0.945
β_{l0}	(t_l)	-0.25464	0.05971	0.000
β_{l1}	($t_l\times$ child is man)	-0.10354	0.02051	0.000
β_{l2}	($t_l\times$ number children)	0.02203	0.00907	0.015
β_{l3}	($t_l\times$ man$\times$number children)	-0.03900	0.01133	0.001
β_{l4}	($t_l\times$ age child)	0.00564	0.00084	0.000
β_{l5}	($t_l\times$ child is married)	0.00757	0.01535	0.622
β_{l6}	($t_l\times$ child is divorced)	-0.01272	0.02226	0.568
β_{l7}	($t_l\times$ child is widowed)	0.04751	0.04445	0.285
β_{l8}	($t_l\times$ child has low education level)	0.13144	0.03673	0.000
β_{l9}	($t_l\times$ child has high education level)	-0.03662	0.01288	0.004
β_{c0}	(c)	0.02788	0.00410	0.000
β_{c1}	($t_c\times$ child is man)	0.00292	0.00161	0.070
β_{c2}	($t_c\times$ number children)	-0.00096	0.00087	0.271
β_{c3}	($t_c\times$ man$\times$number children)	0.00106	0.00092	0.249
β_{c4}	($t_c\times$ age child)	-0.00050	0.00007	0.000
β_{c5}	($t_c\times$ child is married)	0.00348	0.00133	0.009
β_{c6}	($t_c\times$ child is divorced)	0.00355	0.00203	0.080
β_{c7}	($t_c\times$ child is widowed)	0.00235	0.00572	0.682
β_{c8}	($t_c\times$ child has low education level)	0.00288	0.00483	0.552
β_{c9}	($t_c\times$ child has high education level)	0.00320	0.00097	0.001
β_{s0}	(t_s)	-3.03510	0.31385	0.000
β_{s1}	($t_s\times$ child is man)	-0.23573	0.11326	0.037
β_{s2}	($t_s\times$ number children)	0.03988	0.04162	0.338
β_{s3}	($t_s\times$ man$\times$number children)	-0.11870	0.05818	0.041
β_{s4}	($t_s\times$ father good / very good health)	-0.10545	0.18083	0.560
β_{s5}	($t_s\times$ father fair health)	0.88709	0.17274	0.000
β_{s6}	($t_s\times$ father poor health)	1.11773	0.20649	0.000
β_{s7}	($t_s\times$ mother good / very good health)	0.57372	0.15491	0.000
β_{s8}	($t_s\times$ mother fair health)	0.72741	0.14136	0.000
β_{s9}	($t_s\times$ mother poor health)	1.06507	0.16237	0.000
β_{s10}	($t_s\times$ both poor, or poor and fair health)	1.01035	0.16383	0.000
β_{s11}	($t_s\times$ both fair, or fair and good health)	0.43515	0.13891	0.002
β_{s12}	($t_s\times$ father poor, mother good health)	0.67826	0.24108	0.005
β_{s13}	($t_s\times$ father good, mother poor health)	1.36701	0.22906	0.000

[a] Table continues on the next page.

Table 4.5.3: Estimation results structural model, continued[a]

		Coef.	Std. err.	p-value
β_{s14}	($t_s\times$ Germany)	0.34698	0.12951	0.007
β_{s15}	($t_s\times$ Italy)	-0.05540	0.14384	0.700
β_{s16}	($t_s\times$ Greece)	0.41713	0.12893	0.001
β_{s17}	($t_s\times$ Spain)	0.21706	0.14050	0.122
β_{s18}	($t_s\times$ France)	0.22165	0.12188	0.069
β_{s19}	($t_s\times$ Netherlands)	-0.38264	0.18871	0.043
β_{s20}	($t_s\times$ Denmark)	0.18658	0.15070	0.216
β_{s21}	($t_s\times$ Belgium)	0.34925	0.12576	0.005
β_{s22}	($t_s\times$ Austria)	0.27338	0.13071	0.036
β_{s23}	($t_s\times$ Poland)	-0.30224	0.18413	0.101
β_{s24}	($t_s\times$ Czech Republic)	0.59367	0.15322	0.000
β_{s25}	($t_s\times$ (average) age parent - 55)	0.03911	0.00492	0.000
β_{s26}	($t_s\times$ child has low education level)	0.36887	0.11600	0.001
β_{s27}	($t_s\times$ child has high education level)	-0.25675	0.07192	0.000
σ_l^2		0.02078	0.00442	0.000
σ_c^2		0.00007	0.00001	0.000
σ_s^2		1.02037	0.21353	0.000
σ_{lc}		-0.00121	0.00022	0.000
σ_{ls}		0.02571	0.00853	0.003
σ_{cs}		-0.00148	0.00042	0.000
Log likelihood		-2814.912		
N		2253		

[a] The reference individual is a female adult child who has never been married, of whom both parents are alive, have a good / very good health position, and are living in Sweden.

health receive more informal care than mothers in good health (which is also as expected, if fathers in the observed generation indeed have lower household management skills). In addition to poor health, the preference for informal care increases with the age of the parent(s). This is in accordance with the literature, indicating that even after extensively controlling for disability, age remains an important driver of long term care use (De Meijer et al., 2009). The country specific dummy variables comprise institutional as well as cultural differences between countries. Institutional differences constitute for example publicly financed long term care programmes,[14] and the availability of formal care. Cultural differences include differences in social norms with regard to informal care and the degree to which family ties are considered to be important. It has been found that southern European countries have stronger family ties than

[14]An overview of publicly financed long term care programmes can be found in Bolin et al. (2008b).

northern European countries (Reher, 1998). The estimation results show that preferences with regard to informal care are relatively high in Greece, Germany, Belgium, Austria, and the Czech Republic.[15] Higher educated children have significantly lower preferences for informal care than lower educated children. One argument in the literature is that higher educated children live farther away from their parents due to geographical labor market restrictions. However, also after taking into account distance we find a significant effect of education on the preference for informal care, which may be explained by different value orientations of the higher educated (Kalmijn, 2006) and/or competing interests (Waite and Harrison, 1992).[16] Finally, we find that women have significantly higher preferences for providing informal care than men.

Secondly, we describe the parameter estimates related to leisure (t_l). The preference for leisure increases with age and is somewhat lower for men than for women. Children increase women's preferences for leisure significantly, probably because more children often mean more responsibilities for adult daughters inside their own households (the care for a child also belongs to 'leisure time' in this model). Marital status does not affect adult children's preferences for leisure. Married persons spend leisure time with each other, but on the other hand household production is more efficient for couples than for singles, which saves time.[17] Finally, lower educated children have significantly higher preferences for leisure, and higher educated children have significantly lower preferences for leisure. It is possible that less favorable labor conditions among the lower educated bring about higher preferences for leisure time rather than labor time.

The parameter estimates related to consumption (c) show that older children have significantly lower preferences for consumption. In addition, married persons and higher educated individuals have a relatively high preference for consumption. As mentioned before, for the model to be economically rational, the marginal utility of consumption must be positive. We find that for all but

[15] It is notable that southern European countries like Italy and Spain do not have significantly positive results here. Probably this has to do with living arrangements. In Italy and Spain many adult care givers co-reside with their parents and these households are not included in this analysis.

[16] Kalmijn (2006) found that face-to-face contact between higher educated children and their parents is relatively low, even after controlling for distance.

[17] Waite and Harrison (1992) found that the presence of a husband decreases the number of visits a woman has with friends, but does not reduce a woman's social contacts with kin.

18 observations (0.8%) this condition holds. These 18 adult children have a high age, which leads to a relatively low preference for consumption in the model (as can be seen from the coefficient β_{c4}).

The final part of table 4.5.3 shows the estimates of the covariance matrix of the unobserved heterogeneity terms (equation 4.4), which are in line with our expectations. All coefficients are significant, indicating that unobserved heterogeneity is important. The negative sign of σ_{lc} indicates that unobserved characteristics which increase the preference for leisure tend to have a negative effect on the preference for consumption. In the same way, the negative value for σ_{cs} indicates that unobserved characteristics which increase the preference for informal care, have in general a negative effect on the preference for consumption. Finally, σ_{ls} shows that if individuals have a relatively high preference for leisure (conditional on the observed characteristics in the model), they also have on average a somewhat higher preference for informal care.

The relations between wage rates, distances, and informal care follow from the estimated preference parameters and the time and budget constraints. To facilitate interpretation of the results, figure 4.1 shows the relation between geographical distance and the amount of informal care given by a reference individual in the model. As a reference individual we consider a married German woman of age 55, with an 80 year old father in poor health, no mother, and 2 children of her own. She has a medium education level, a wage rate of 10 euros per hour and her remaining household income is 15,000 euros per year. Unobserved heterogeneity is important regarding the preferences for informal care. Figure 4.1 therefore shows seven lines. Each line represents the reference individual with a different random effect u_s. These reflect, for example, different levels of family ties, degree of altruism, or feelings of obligation to provide informal care. The line 'p50' shows the relationship between distance and informal care when all random effects u_l, u_c and u_s are equal to zero. This means that the unobserved preferences with regard to leisure, consumption, and informal care are at the median level. For example, with regard to informal care we can interpret this reference individual to have 'median responsibility norms'. The line 'p90' represents the reference individual with high unobserved preferences for informal care. Only 10% of the individuals have a higher random effect u_s. The same explanation holds for the other lines, p10, p25, p60, p70, and p80. For example, for line p25, only

25% of the people have smaller unobserved preferences for informal care. u_l and u_c are zero for all lines, such that the only difference between the lines is the random effect u_s, the unobserved heterogeneity with regard to informal care.

Figure 4.1: Estimated relationship between distance and the expected supply of informal care to elderly parents for the reference individual

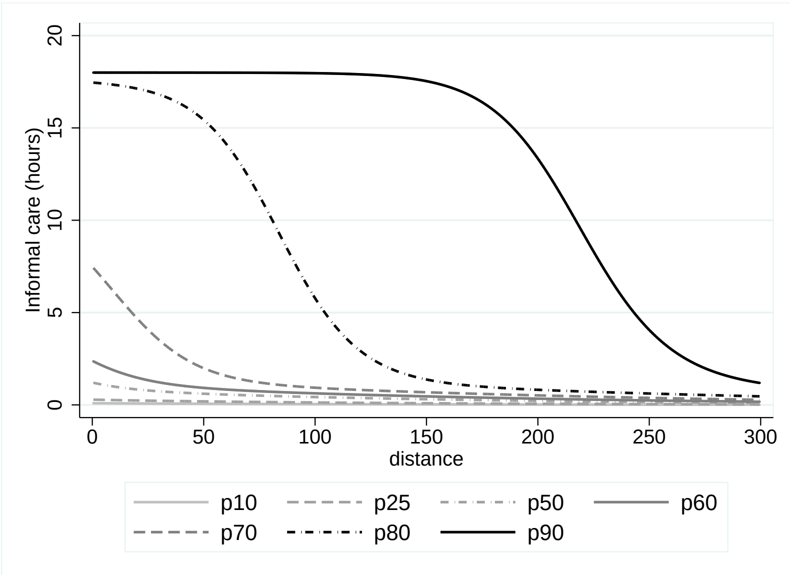


Figure 4.1 shows that the reference person with 'median' unobserved preferences for informal care provides almost no informal care. This is as expected, since we found in table 4.4.3 that only 34% of the only children with a father in poor health provide informal care. The higher the preference of the reference individual to provide care, the longer it takes before informal care decreases with distance (distance elasticity is low for those with high preferences for informal care).

The distance between adult children and their parents may also influence the labor force participation of the adult children. Unsurprisingly, figure 4.2 shows that for the majority of adult children, who give almost no informal care, distance does not influence labor force participation. Focussing on p70, we see that labor supply increases with distance. Apparently, at least part of the reduction in informal care is replaced by labor. For those with relatively high preferences for informal care, labor force participation first declines when distance increases, as more travel time is needed for the provision of informal care. However, after a certain distance (e.g. 50 kilometers for the 80th percentile), informal care decreases and labor force participation increases.

Figure 4.2: Estimated relationship between distance and the expected supply of labor

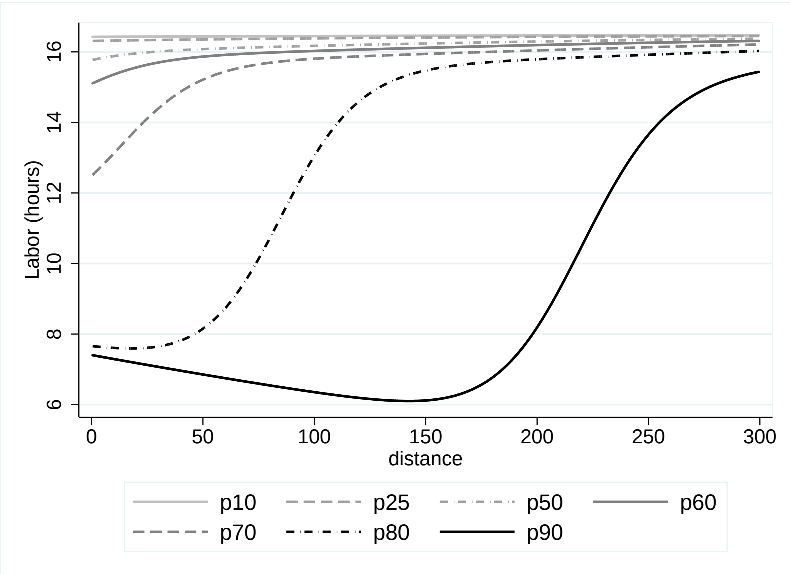


Figure 4.3 and 4.4 show the relation between the wage rate of the reference individual, the expected number of hours of participation in the labor market, and the hours of informal care the reference individual provides to her father. In these figures the distance between the reference individual and her father is 7.5 kilometers. The seven lines represent different levels of the unobserved heterogeneity term with regard to informal care, just as explained for figure 4.1. In line with the literature (e.g. Evers et al., 2008), figure 4.3 shows a positive wage elasticity of labor supply. Reference individuals with higher preferences for informal care are less active in the labor market. For example, at the wage rate of 10 euros per hour, the reference individual with a high preference for informal care (p90) participates about 9 hours less in the labor market than the reference individual with a low unobserved preference for informal care (p10).

According to Figure 4.4 the wage elasticity of informal care supply is small. The wage elasticity for a reference individual with large norms of responsibility (or other reasons that lead to a high unobserved preference for informal care) is almost zero.

Figure 4.3: Relationship between wages and the expected hours of labor supply for the reference individual

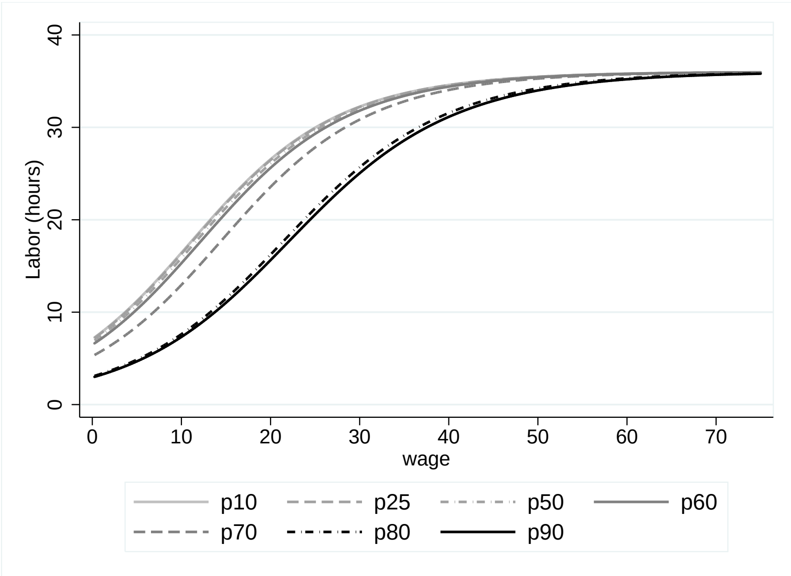
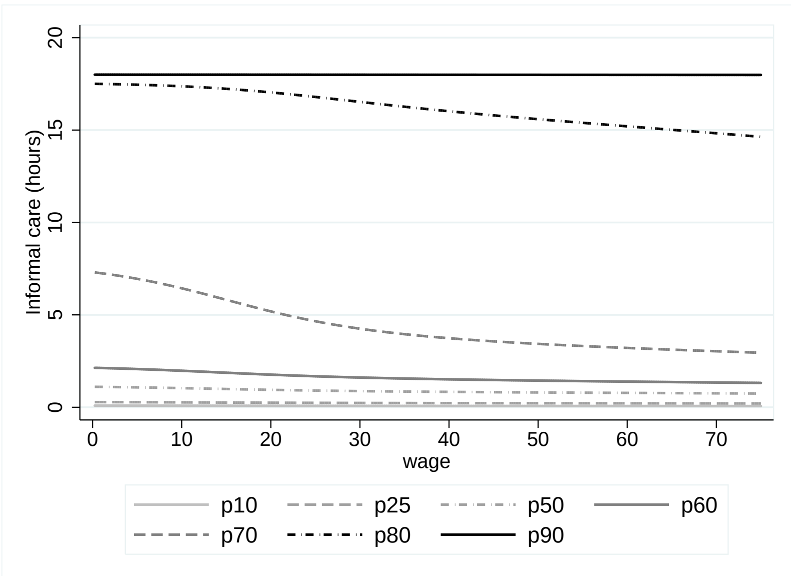


Figure 4.4: Relationship between wages and the expected supply of informal care to elderly parents for the reference individual



Two adult children 4.6

In families with two siblings, informal care provision to parents is determined by the characteristics of both siblings, and the nature of the interaction between siblings. In this section we use the estimates of the structural model, estimated for only children, to families with two adult children. When applying the estimates of only children to siblings, some assumptions are required. First of all, we assume that siblings have the same preferences for leisure, consumption and informal care as only children.[18] The only difference is that there is now a sibling available who can also provide informal care (the hours of informal care t_s in the utility function becomes the sum of own informal care and informal care provided by the sibling). We assume that informal care provided by oneself and by the sibling are perfect substitutes. This means that children receive the same direct utility from an hour of informal care provided by themselves or by their sibling (this utility is β_s, from equation 4.3). Also, an hour of informal care provided by one of the siblings decreases the marginal utility of an extra hour of informal care by α_{ss} (as in the model for only children), it increases the utility obtained from leisure by α_{ls}, and it changes the marginal utility of consumption by α_{cs} (not significant). Only, for those siblings with a negative direct utility from informal care ($\beta_s < 0$), we assume that they do not receive any direct utility from an extra hour of informal care provided by their sibling (these are, for example, individuals with healthy parents and/or low unobserved preferences for informal care). Finally, we assume that both siblings have their own time and budget constraints and that there are no financial transfers between siblings. The amount of informal care provided may be the outcome of a non-cooperative or cooperative game between two siblings.

Section 4.6.1 describes how we derive non-cooperative and cooperative equilibria. Next, we show some simulations of cooperative and non-cooperative behavior between reference siblings (4.6.2). Finally, we apply the model estimated in this chapter to the families with two siblings in SHARE, to gain an indication of whether siblings behave cooperatively or non-cooperatively, and to estimate the expected gains from cooperation between siblings (4.6.3).

[18]We see no reason why preferences for leisure, consumption, and informal care would be different for only children, compared to children with brothers and sisters. Regarding informal care, this is in line with Spitze and Logan (1991), who find that the number of siblings is unrelated to adult children's closeness to parents or attitudes toward filial responsibility.

4.6.1 Cooperative and non-cooperative equilibria

Non-cooperative equilibrium

In the non-cooperative equilibrium, we assume that both siblings maximize their utility, given the choice of their sibling and their own time and budget constraints. We use a generalization of the Nash equilibrium, based on the assumption that a player's rationality is bounded. Bounded rationality is incorporated by adding random disturbance to the payoffs of the players, just as we did for only children in (4.5). Just as for only children, we assume that siblings are more likely to choose better strategies than worse strategies, but do not play the best strategy with probability one (children are 'better responders' rather than 'best responders'). This concept, in a game-theoretic framework, has been explained by McKelvey and Palfrey (1995, 1998) and is called the Quantal Response Equilibrium (QRE). As we add random errors distributed according to the type I extreme value distribution, we have a special version of the Quantal Response Equilibrium, namely the logit equilibrium (LQRE). The LQRE extends the model we estimated for only children to the situation with two or more siblings. In the logit equilibrium the sibling's alternatives are chosen according to the probability distribution

$$p_{i,m} = \frac{\exp\left(\lambda E(U(t_{i,m} \mid p_j))\right)}{\sum_{k=1}^{12} \exp\left(\lambda E(U(t_{i,k} \mid p_j))\right)} \quad m = 1, \ldots, 12 \tag{4.19}$$

where p_{im} is the probability of sibling i choosing alternative m. $E\left(U(t_{i,m} \mid p_j)\right)$ is the expected utility to player i of choosing alternative m when sibling j has probability distribution p_j for the 12 alternatives. The time and budget constraints are substituted in the utility function. The nonnegative parameter λ is inversely related to the level of error and can be interpreted to reflect the degree of bounded rationality. When $\lambda \to \infty$, players become 'perfectly rational' and the logit equilibrium converges to the Nash equilibrium. In the other extreme case, when $\lambda = 0$, the probabilities of the twelve alternatives converge to $1/12$, for both siblings (i.e., siblings make extremely noisy choices). Standard multinomial logit models assume $\lambda = 1$. Consistent with the model for only children, we also assume $\lambda = 1$.[19]

[19] The parameter λ can not be identified. For future research it may be interesting to allow λ to vary with the education levels of the adult children.

The logit response functions p_i and p_j are functions of each other. For example, the probability of sibling 1 choosing alternative m depends on the probabilities of sibling 2 choosing alternatives 1 to 12. On the other hand, the probability of sibling 2 choosing alternative m depends on the probabilities of sibling 1 choosing alternatives 1 to 12. We find the logit equilibrium by solving the logit response functions, which form a system of 24 nonlinear equations that are listed in Appendix 4.6.2.

Cooperative equilibrium

In the cooperative equilibrium, we assume that siblings maximize the sum of their utilities.

$$U\left(t_1, t_2\right) = \gamma\, U\left(t_1\right) + (1-\gamma)\, U\left(t_2\right) \quad \gamma \in [0,1] \tag{4.20}$$

subject to their own time and budget constraints. We choose $\gamma = 0.5$, which is one choice out of the large set of Pareto solutions.[20]

For each of the $12 \times 12 = 144$ possible alternatives for the two siblings we compute $U(t_1, t_2)$, and we use these utilities to compute the probability of each alternative in the same way as we did for only children (equation 4.6). The probability of alternative l is

$$q_l = \frac{\exp\left(U(t_{1,l}, t_{2,l})\right)}{\sum_{k=1}^{144} \exp\left(U(t_{1,k}, t_{2,k})\right)} \quad l = 1, \ldots, 144. \tag{4.21}$$

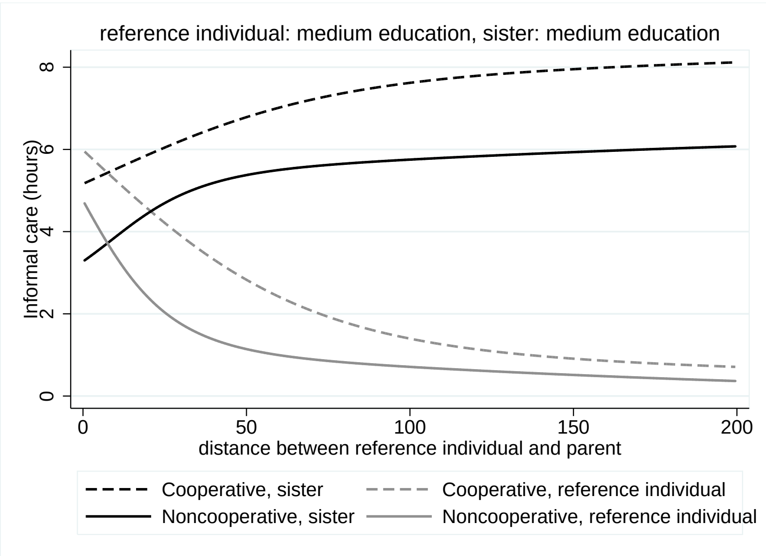
Simulations 4.6.2

In this section we simulate some non-cooperative and cooperative equilibria. For these simulations we stick to our reference person, specified in section 4.5.2 (a woman of age 55, who is married and living in Germany, who has an 80 year old father in poor health, no mother, two children, a medium education level, a remaining household income of 15,000 euros per year, and a wage

[20] In section 4.6.3 we do a sensitivity analysis, which shows that the conclusions are not very sensitive to the choice of the weights.

rate of 10 euros per hour). However, in this section our reference person is no longer an only child. First, we assume that she has a sister, who has exactly the same characteristics as herself. This sister lives 7.5 kilometers from the parent. Figure 4.5 presents the amount of care that these two sisters are providing to their parent, for different geographical distances of our reference individual.[21] At the distance of 7.5 kilometers, both siblings have exactly the same characteristics, and we see that they indeed give the same amount of care. When the distance of our reference individual to the parent increases, the amount of care provided by our reference sibling decreases, but the amount of care provided by her sister increases (she compensates for part of the loss of informal care). The total amount of informal care provided is higher in the cooperative equilibria than in the non-cooperative equilibria. Compared to the situation where our reference person was an only child (p70 in figure 4.1), in the non-cooperative equilibria our reference person provides substantially less informal care.

Figure 4.5: Cooperative and non-cooperative outcomes for two siblings, by distance of the reference individual



In figure 4.6 the sister of our reference person, explained above, has a high education level instead of a medium education level (the two sisters still

[21]In section 4.5.2 we found that only those who have a relatively high unobserved random effect for informal care provide informal care. In figure 4.5 we therefore assume the sisters to have unobserved preferences for informal care at the 70th percentile, corresponding to the line 'p70' in section 4.5.2 (30% of the of the adult children have a higher random effect u_s).

have the same wage rate). In the presence of the higher educated sister, our reference person provides more informal care in the non-cooperative equilibria than in the cooperative equilibria (until about 40 kilometers), because she has a higher preference for informal care than her sister. When the reference

Figure 4.6: Cooperative and non-cooperative outcomes for two siblings, by distance of the reference individual

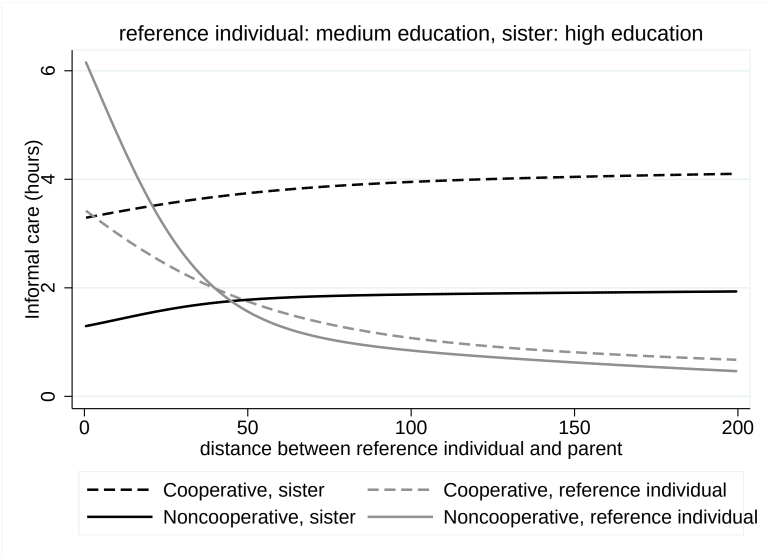
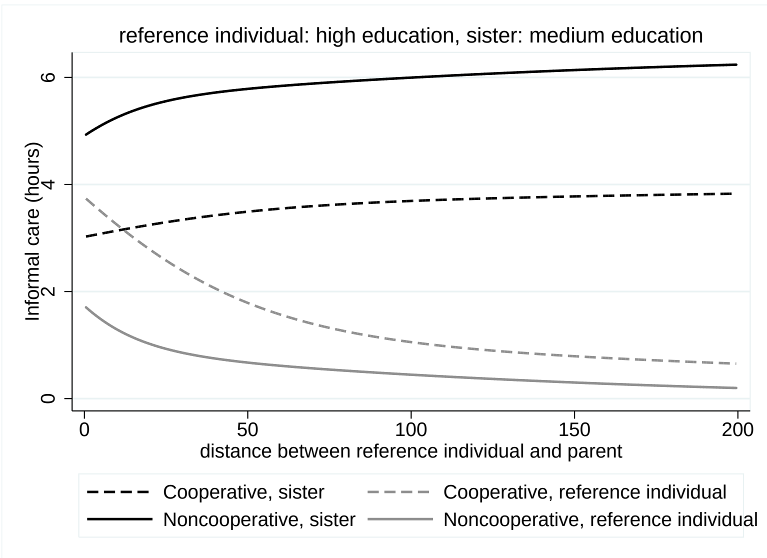


Figure 4.7: Cooperative and noncooperative outcomes for two siblings, by distance of sibling 2



individual lives farther from their father, her higher educated sister increases her provision of informal care slightly.

When we switch the education levels for our reference person and her sister (such that the reference person has a high education level and her sister has a medium education level), we find the equilibria shown in figure 4.7. In the non-cooperative equilibria the medium educated sister provides most of the informal care, whereas in the cooperative equilibria informal care is more shared between the reference individual and her sisters.

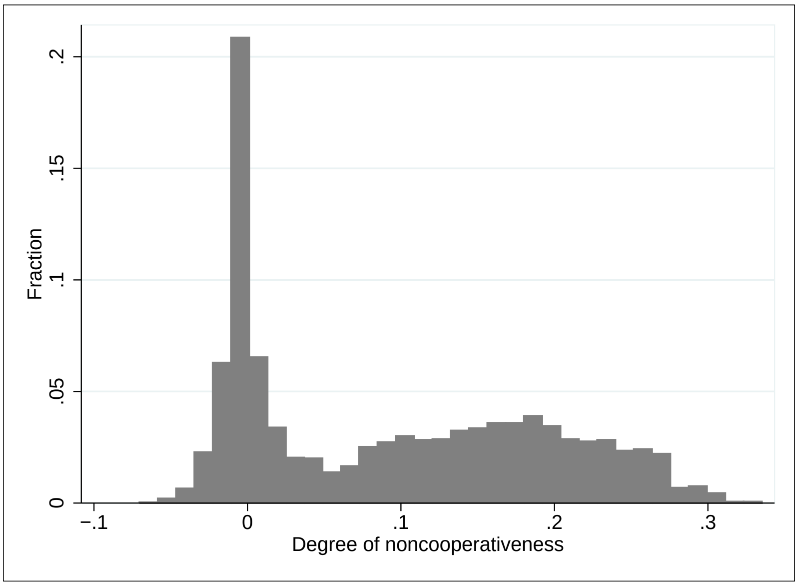
4.6.3 Interactions between siblings in SHARE

The simulations in the previous section showed us that the nature of the interactions between children can have a large effect on the division of informal care between siblings and the total amount of informal care provided to parents. In this section we apply the estimated structural model to families with two adult children in SHARE (described in section 4.4). First, we examine the fit of cooperative and non-cooperative equilibria. Second, we investigate which siblings behave cooperatively and non-cooperatively (using observed characteristics). Finally, we study the gains that can be achieved by cooperation.

To examine the fit of cooperative and non-cooperative equilibria, we predict the cooperative and non-cooperative outcomes for the siblings in SHARE (using their observed characteristics and the structural parameter estimates from the only child empirical results), and compare them with their realized outcomes. Cooperative and non-cooperative equilibria are described by probabilities for each of the twelve alternatives described in section 4.3, for both siblings. We examine the fit of the cooperative and non-cooperative equilibria by the sum (over siblings) of the probabilities for the realized options, divided by the number of siblings. This can be interpreted as the percentage of correct predictions of the model. The non-cooperative model has a higher fit than the cooperative model (26.8% versus 17.3%). In the non-cooperative model siblings provide on average 1.13 hours of informal care per week, whereas in the cooperative model this is 1.63 hours. The realized average hours of informal care is also closer to the non-cooperative outcome than to the cooperative outcome, namely 1.18 hours per week.

The next question we want to answer is which people tend to behave cooperatively and which people tend to behave non-cooperatively. We measure the degree of non-cooperativeness by the difference between the non-cooperative and the cooperative predicted probabilities for the realized outcome. Figure 4.8 shows the histogram of this measure of non-cooperativeness and can be interpreted as follows: when the degree of non-cooperativeness is 0.1, the realized outcome has a 10%-points higher probability to be a non-cooperative than a cooperative outcome. The histogram shows that most of the families (71%) have a higher probability to behave non-cooperatively than cooperatively. Even, for 47% of the families the probability that they behave non-cooperatively is 10%-points higher than the probability that they behave cooperatively.[22] The spike around zero includes families for whom the cooperative and the non-cooperative outcomes are about the same.

Figure 4.8: Histogram degree of non-cooperativeness



We regress the degree of non-cooperativeness on several background characteristics of the siblings. The results in table 4.6.1 show that, relative to two

[22]A sensitivity analysis shows that the conclusions are not very sensitive to the weights chosen in equation (4.20). For example, when we choose the weights to be 0.3 and 0.7 (instead of 0.5 and 0.5) the fit of the cooperative model is 14.6% instead of 17.3%, the number of hours of informal care is 1.81 instead of 1.63, 75% of the families have a higher probability for the non-cooperative equilibrium (instead of 71%), and 50% of the families have a 10%-points higher probability to behave non-cooperatively instead of cooperatively (instead of 47% when the weights are 0.5 and 0.5).

sisters, two brothers have on average a 10.5%-points higher probability to behave non-cooperatively than cooperatively. Also, a brother-sister relationship appears to be more cooperative than a brother-brother relationship. This may be explained by the fact that traditionally women are kin keepers. It has been found that sister-to-sister relationships and sister-to-brother relationships show on average greater emotional closeness and more frequency of contact than brother-brother relationships (Connidis and Campbell, 1995). Furthermore, when both of the siblings have a high education level, or when one of them has a high education level and the other a medium education level, they are significantly less cooperative than two medium or low educated siblings. On average, two high educated siblings have a 3.2%-points higher probability to behave non-cooperatively rather than cooperatively. Finally, older siblings and siblings with a larger age difference have a significantly higher probability to behave cooperatively, and the differences in cooperativeness between countries is small. Siblings in Austria and the Netherlands behave slightly more cooperatively than siblings in Sweden, while siblings in Spain, Italy and Denmark behave somewhat less cooperatively.

To obtain insights into the gains that can be achieved from cooperation, we compute the increase in the hours of informal care that would occur if those who seem to be non-cooperative were to change to cooperative behavior. If those who tend to be non-cooperative (who have a higher probability to be non-cooperative than to be cooperative) were pushed into cooperative behavior, their average provision of informal care would increase from 1.04 hours per week to 1.52 hours per week. So, their parents would on average receive 0.96 hours of informal care per week more from their children, which is a growth of 46.2%. While informal care increases when families are pushed into their cooperative outcome, the number of individuals working full-time in the labor market decreases by 5.7%-points and the number of individuals working part-time increases by 6.7%-points.

Table 4.6.1: Degree of non-cooperativeness

	Coef.	Std. err.	p-value
Gender			
2 sisters	0.000	-	-
brother and sister	0.020	0.004	0.00
2 brothers	0.105	0.005	0.00
Age			
Age youngest sibling	-0.004	0.000	0.00
Age difference between the siblings	-0.002	0.001	0.00
Education			
Both low education level	-0.014	0.010	0.17
Both high education level	0.032	0.004	0.00
Low and medium education level	-0.018	0.008	0.02
Low and high education level	-0.006	0.015	0.67
Medium and high education level	0.021	0.004	0.00
Number of children			
Minimum number of children of both siblings	-0.004	0.002	0.07
Difference in number of children between siblings	0.000	0.002	0.91
Partners			
No partners	0.000	-	-
One sibling has a partner	-0.004	0.009	0.63
Both siblings have a partner	-0.006	0.009	0.53
Country of the parents			
Sweden	0.000	-	-
Austria	-0.016	0.007	0.04
Belgium	0.001	0.007	0.85
Germany	-0.012	0.007	0.08
Denmark	0.019	0.007	0.01
Spain	0.031	0.008	0.00
France	0.004	0.007	0.54
Italy	0.030	0.008	0.00
The Netherlands	-0.019	0.007	0.01
Czech Republic	-0.011	0.007	0.12
Greece	-0.005	0.006	0.45
Poland	0.013	0.010	0.16
Constant	0.238	0.017	0.00
N	2891		
R^2	0.251		

4.7 Conclusions

This chapter presents a structural model to analyze families' complex decisions regarding informal care provision for aging parents. In the model adult children maximize their utility, defined over consumption, leisure, and the amount of care that parents receive from their children, subject to a time and budget constraint.

In the first part of this study, the preference parameters of the model are estimated using only children, such that interactions between siblings do not play a role. The results show that the preference for informal care is influenced by the health of the parents, the gender and education level of the adult children, and cultural and institutional differences between countries. Also unobserved individual specific preferences such as altruism, reciprocity and responsibility norms play a large role in the preferences of adult children to give informal care. The (negative) wage elasticity of informal care supply appears to be small.

The second part of the chapter focuses on the strategic interactions between siblings. In the literature it has been emphasized that modeling family decisions as a bargaining process is important to increase our understanding of these decisions. An important follow-up question is whether this bargaining process is cooperative or non-cooperative. In a structural model with two siblings one has to make assumptions about the nature of the interactions between siblings. When some families are cooperative and other non-cooperative, this cannot be identified in general together with the other coefficients in a game-theoretic model. In some way, one needs information about the (non-)cooperativeness of siblings, which is often not available. Most often, empirical game-theoretic models assume that siblings make non-cooperative decisions. This study presents a first attempt to identify the nature of the interactions between siblings using the structural parameter estimates of only children. We show that the nature of the interactions between siblings can have a large effect on the division of informal care between siblings and the total amount of informal care provided to parents. Furthermore, it appears that 71% of the siblings have a higher probability to be non-cooperative than cooperative (which means that the assumption of non-cooperative siblings used by current game-theoretic models holds for the majority of the siblings). The degree of cooperativeness varies most with the gender of the siblings. On average, two brothers have a 10.5% higher probability to be non-cooperative than two sisters. Furthermore, two higher educated siblings or a higher and medium educated sibling appear to be less cooperative on average than two medium or lower educated siblings, and older siblings have a significantly higher probability to behave cooperatively.

For future research it may be interesting to estimate this model using U.S. data from the Health and Retirement Study (HRS). The HRS has the advantage that it also contains information about family income of the adult children.

For policy design, we can conclude that a reduction in the geographical distance between adult children and their parents would be an effective measure to increase informal care as well as the labor force participation of those children with a relatively high preference for informal care. For example, the social rent sector could weigh informal care in their assignment of houses, or senior houses could be built in residential areas. For fiscal policies it may be of interest that net wages have negligible effects on the provision of informal care, while they do influence labor supply. Pushing non-cooperative families into their cooperative equilibria would increase the provision of informal care, but this would be at the expense of the labor supply of adult children.

4.A Wage equations per country

Table 4.A.1: Wage equations per country

Equation 1: ln(wage rate)	AT	BE	DE
Man	0.095	-0.016	0.062**
Age	0.066***	0.044***	0.098***
$Age^2/100$	-0.067***	-0.042***	-0.100***
Primary education	-1.511***	-0.095	-0.590***
Secondary education	-0.261***	-0.140***	-0.287***
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.	0.155**	-0.006	0.125***
Tertiary education	0.388***	0.178***	0.270***
Man with partner	0.045	0.064	0.193***
Woman with partner	-0.042	-0.052	-0.056**
Intercept	0.574**	1.243***	-0.217*
Equation 2: participation decision			
Man	0.023	0.453**	-0.283***
Age 15–29	0.690***	1.197***	0.694***
Age 30–39	0.793***	1.160***	0.860***
Age 40–49	-0.393***	0.260*	0.343***
Age $\geq$ 60	-2.997***	-2.376***	-1.988***
Primary education	-1.132**	-0.740***	-1.036***
Secondary education	-0.288***	-0.474***	-0.330***
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.	0.015	0.459	0.120
Tertiary education	0.382***	0.634***	0.269***
Man with partner	0.738***	0.971***	0.704***
Woman with partner	-0.078	0.363***	-0.130**
Man with child	-0.195	-0.494***	-0.021
Woman with child	-0.827***	-0.417***	-0.695***
Intercept	0.510***	-0.627***	0.114
ρ	0.083	-0.116	0.196***
σ_w	0.490***	0.323***	0.409***
$\sigma_{wp} = \rho\sigma_w$	0.041	-0.037	0.080***
Observations	1488	1346	6028
Censored observations	658	676	3240
Uncensored observations	830	670	2788
Log likelihood	-1165.866	-704.823	-4038.892

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4.A.1: Wage equations per country (*continued)*

Equation 1: ln(wage rate)	DK	ES	FR
Man	-0.114	0.147***	0.080*
Age	0.086***	0.031***	0.070***
$\text{Age}^2/100$	-0.090***	-0.025***	-0.066***
Primary education		-0.262***	-0.204***
Secondary education	-0.136***	-0.163***	-0.042
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.		-0.081	0.378***
Tertiary education	0.158***	0.294***	0.407***
Man with partner	0.297***	0.137***	0.062
Woman with partner	0.006	0.088***	-0.035
Intercept	0.474	0.932***	0.423**
Equation 2: participation decision			
Man	0.231	0.190***	-0.117
Age 15–29	1.335***	0.748***	1.115***
Age 30–39	1.556***	0.650***	1.224***
Age 40–49	1.321***	-0.053	0.397***
Age $\geq$ 60	-1.339***	-1.942***	-2.587***
Primary education		-0.259***	-0.317***
Secondary education	-0.550***	-0.029	-0.405***
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.		0.495***	0.575**
Tertiary education	0.198*	0.623***	0.475***
Man with partner	0.628***	0.923***	1.138***
Woman with partner	0.520***	-0.154***	0.399***
Man with child	-0.057	-0.287***	-0.349***
Woman with child	-0.222	-0.547***	-0.697***
Intercept	-0.410***	0.031	-0.039
ρ	0.040	-0.028	0.035
σ_w	0.515***	0.452***	0.466***
$\sigma_{wp} = \rho\sigma_w$	0.021	-0.013	0.016
Observations	1422	7171	3221
Censored observations	509	3328	1321
Uncensored observations	913	3843	1900
Log likelihood	-1160.870	-5647.577	-2396.299

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4.A.1: Wage equations per country (*continued)*

Equation 1: ln(wage rate)	CZ	GR	PL
Man	0.141**	0.135**	0.006
Age	0.026**	0.070***	0.037***
$\text{Age}^2/100$	-0.032**	-0.065***	-0.030**
Primary education		-0.226***	-0.299***
Secondary education	-0.207***	-0.051	-0.810***
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.	0.279*	0.233***	0.208***
Tertiary education	0.398***	0.353***	0.608***
Man with partner	0.195***	0.148***	0.154***
Woman with partner	0.039	0.144***	0.025
Intercept	-0.055	-0.095	-0.698***
Equation 2: participation decision			
Man	-0.278	0.334**	-0.053
Age 15–29	0.834***	1.070***	0.729***
Age 30–39	1.447***	0.717***	0.688***
Age 40–49	-0.080	-0.207	-0.244***
Age $\geq$ 60	-3.139***	-2.553***	-2.027***
Primary education		-0.345***	-0.632***
Secondary education	-1.248***	-0.444***	-1.734***
(Upper) secondary education	0.000	0.000	0.000
Post secondary non-tertiary edu.	0.941	0.784***	0.443***
Tertiary education	0.501**	0.591***	0.797***
Man with partner	0.990***	1.364***	0.935***
Woman with partner	-0.109	-0.010	0.259***
Man with child	-0.978***	-0.547***	-0.050
Woman with child	-1.320***	-0.263**	-0.290***
Intercept	1.124***	-0.411***	-0.324***
ρ	0.241*	-0.075	-0.110
σ_w	0.375***	0.388***	0.543***
$\sigma_{wp} = \rho\sigma_w$	0.091*	-0.029	-0.060
Observations	1095	1345	10464
Censored observations	485	729	5883
Uncensored observations	610	616	4581
Log likelihood	-584.078	-832.007	-8276.916

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4.A.1: Wage equations per country (*continued)*

Equation 1: ln(wage rate)	IT	NL
Man	0.095***	-0.042*
Age	0.045***	0.063***
$Age^2/100$	-0.040***	-0.065***
Primary education	-0.310***	-0.170***
Secondary education	-0.169***	-0.147***
(Upper) secondary education	0.000	0.000
Post secondary non-tertiary edu.	0.040**	0.046*
Tertiary education	0.301***	0.237***
Man with partner	0.090***	0.222***
Woman with partner	0.067***	-0.048**
Intercept	0.939***	0.876***
Equation 2: participation decision		
Man	0.223***	0.018
Age 15–29	0.914***	0.702***
Age 30–39	1.085***	0.662***
Age 40–49	0.078*	-0.261***
Age $\geq$ 60	-1.945***	-2.883***
Primary education	-0.798***	-0.508***
Secondary education	-0.395***	-0.489***
(Upper) secondary education	0.000	0.000
Post secondary non-tertiary edu.	0.380***	0.044
Tertiary education	0.400***	0.356***
Man with partner	0.712***	1.144***
Woman with partner	-0.078*	0.176***
Man with child	-0.305***	-0.338***
Woman with child	-0.628***	-0.706***
Intercept	-0.018	0.499***
ρ	0.057	0.133**
σ_w	0.382***	0.319***
$\sigma_{wp} = \rho\sigma_w$	0.022	0.042**
Observations	14155	6007
Censored observations	7740	2195
Uncensored observations	6415	3812
Log likelihood	-8780.796	-3100.034

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.B Logit equilibrium

Section 4.6.1 explains the non-cooperative logit equilibrium, which is a generalization of the Nash equilibrium and deals with 'noisy decisions' made by bounded-rational siblings. This equilibrium concept extends the model for only children described in section 4.3, to a game theoretic framework with two players.

In section 4.6 we have two siblings, i and j who can choose between 12 alternatives. Therefore, to obtain the logit equilibrium we have to solve a system of 24 nonlinear equations, the logit response functions. The logit response functions of sibling i are

$$
\begin{aligned}
p_{i,1} &= \frac{\exp\left(U(t_{i,1} \mid j=1)p_{j,1} + U(t_{i,1} \mid j=2)p_{j,2} + \cdots + U(t_{i,1} \mid j=12)p_{j,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{i,k} \mid j=m)p_{j,m}\right)} \\
p_{i,2} &= \frac{\exp\left(U(t_{i,2} \mid j=1)p_{j,1} + U(t_{i,2} \mid j=2)p_{j,2} + \cdots + U(t_{i,2} \mid j=12)p_{j,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{i,k} \mid j=m)p_{j,m}\right)} \\
&\vdots \\
p_{i,12} &= \frac{\exp\left(U(t_{i,12} \mid j=1)p_{j,1} + U(t_{i,12} \mid j=2)p_{j,2} + \cdots + U(t_{i,12} \mid j=12)p_{j,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{i,k} \mid j=m)p_{j,m}\right)}
\end{aligned}
$$

The logit response functions of sibling j are

$$
\begin{aligned}
p_{j,1} &= \frac{\exp\left(U(t_{j,1} \mid i=1)p_{i,1} + U(t_{j,1} \mid i=2)p_{i,2} + \cdots + U(t_{j,1} \mid i=12)p_{i,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{j,k} \mid i=m)p_{i,m}\right)} \\
p_{j,2} &= \frac{\exp\left(U(t_{j,2} \mid i=1)p_{i,1} + U(t_{j,2} \mid i=2)p_{i,2} + \cdots + U(t_{j,2} \mid i=12)p_{i,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{j,k} \mid i=m)p_{i,m}\right)} \\
&\vdots \\
p_{j,12} &= \frac{\exp\left(U(t_{j,12} \mid i=1)p_{i,1} + U(t_{j,12} \mid i=2)p_{i,2} + \cdots + U(t_{j,12} \mid i=12)p_{i,12}\right)}{\sum_{k=1}^{12} \exp\left(\sum_{m=1}^{12} U(t_{j,k} \mid i=m)p_{i,m}\right)}
\end{aligned}
$$

These 24 equilibrium conditions have to be solved numerically since there is no closed-form solution.

5 Changes in the Income Distribution of the Dutch Elderly between 1989 and 2020: A Dynamic Microsimulation

This chapter is based on Knoef, Alessie, and Kalwij (2009).

5.1 Introduction

In 2011, the first generation of the babyboom will reach the statutory retirement age of 65. From then onwards, there will be a doubling in the proportion of retirees over the working population from 26% in 2011 to 47% in 2038. This places an increasing financial burden on society through pay-as-you-go financed social security, pension, health, and long-term care systems.

Policies aimed at alleviating the costs related to the aging society can be based on the notion that the financial burden is shared between generations (see Bovenberg and Ter Rele, 2000, Van Ewijk et al., 2006). Alternatively or at the same time, one could call upon intragenerational solidarity, such as solidarity within the elderly generations. An example of a proposed policy is lowering the indexation ambition of the public pensions (Den Butter, 2010). Another example is 'fiscalization' of the public pension contributions. In this case, a larger part of the pay-as-you-go public pension scheme will be financed by general tax revenues. Consequently, also the 65+ population pays for the state pensions and due to the progressive Dutch tax system, this policy option redistributes income within the elderly generation.

In order to assess the viability of proposed reforms, policymakers require insights into the income distribution of current and future generations of pensioners in a situation of no policy changes. It is important to note that also without pension reforms, the future income distribution of pensioners will differ from the current distribution due to developments in longevity and in demographic and socio-economic compositions. For instance, the number of divorces is increasing and female labor force participation has increased strongly during the last decades, so that many more women will receive an occupational pension income in the future. Also, there are productivity differences between cohorts that lead to income differences.

The contribution of this study to the literature is threefold. First, we describe developments in the income distribution for the age groups 50–64 and 65–90 between 1989–2007 for the Netherlands. We also present developments in the income composition for different parts of the distribution. The data show that occupational pensions have become more important over the whole income distribution, not just at the upper part.

Second, we predict the income distribution of the elderly until 2020 using a microsimulation model. Previous research making predictions about the income distribution in the Netherlands was performed by Dessens and Jansen (1997) and SZW (2006). Dessens and Jansen examined the consequences of the increased proportion of working female partners on trends in income inequality. They extrapolated the Gini coefficient until 2011, using predictions of female participation rates and the average ratio of their incomes to those of their partners. Our empirical results indicate that a single inequality index such as the Gini coefficient is not always sufficiently informative to describe trends in income inequality, as trends in the lower and upper segments of the distribution may be contradictory.

With a microsimulation model detailed estimates on the whole income distribution are possible.[1] SZW (2006) uses a microsimulation model to predict the future income distribution of pensioners in the Netherlands. Our model deviates from this model by using longitudinal instead of cross sectional data, such that we can disentangle age, period, and cohort effects. Furthermore, we use administrative instead of survey data. A related microsimulation study of Van Sonsbeek (2009) predicts the costs and the redistributive effect of public pensions in the Netherlands.

Third, new in this microsimulation study on income is that we investigate

[1]Merz (1991), O'Donoghue (2001) and Zaidi and Rake (2001) explain, review, and classify microsimulation models around the world.

the income process by explicitly paying attention to the modeling of the error terms. Households may experience income shocks, the distribution of which may be different for different types of households. In addition, income shocks may have persistent effects, and the degree of persistency probably increases with age. Therefore, we take into account autocorrelation and allow for the fact that the degree of autocorrelation differs over the lifecycle. We are not aware of a previous microsimulation study on income that takes into account heteroskedasticity and persistency of income shocks.

For the income predictions we estimate a fixed effects income equation with three specifications. The first specification only contains age and period effects. It models no other underlying processes that influence income (for example labor market positions) and thus relies heavily on the modeling of the income process. In the second specification household demographics are added, and the third specification also incorporates the labor market status of household members. In these specifications changes in demographic variables or labor market status lead to income shocks. The main results of the three specifications are rather similar. From this founding we cautiously conclude that adding other background characteristics will not affect the simulation results dramatically.

The advantage of using fixed effects and modeling the error terms is that they make the explicit modeling of underlying processes influencing household income less necessary. Yet, more complex simulation models give more underlying information. For example, only after explicitly modeling labor market status (which we do in specification three), we can say more about the income positions of elderly with and without occupational pension income.

The results show that next generations of pensioners have higher equivalized household incomes than current generations of pensioners, especially among households with median income. Between 2008 and 2020, equivalized household income of the elderly in the age group 65–90 increases on average by 0.5% per year for the 10th percentile, 1.2% for the median, and 1.0% for the 90th percentile. Inequality among pensioners increases at the lower end of the income distribution, but decreases at the upper end. The increased inequality in the lower segment is not the result of a higher inequality between households with and without occupational pension income. Instead, inequality between households with and without occupational pension income in the lower segment of the distribution decreases until 2020.

If one aims to quantify the effects of different pension policies, then it is important to model labor supply responses explicitly (Creedy and Duncan, 2002). This is beyond the scope of this study, however. This study offers insights into the development of the future income distribution, induced by increased longevity and ongoing demographic and socio-economic changes. If labor market outcomes of a certain policy measure are known, they can be incorporated into the model.

This chapter is structured as follows: the next section reviews the empirical literature on developments in the Dutch income distribution in the past. Sections 5.3 and 5.4 describe the data and the microsimulation model, after which section 5.5 summarizes the estimation results. Section 5.6 presents the results of the simulation, and the chapter concludes with section 5.7.

5.2 Developments in the Dutch income distribution

This section shortly reviews the empirical literature on the Dutch income distribution during the past fifty years. From the 1960s onward, inequality decreased rapidly. The main reason for this was the relative increase of the income of inactive households, due to the establishment of the social security system (Caminada and Goudswaard, 2003, Trimp, 2000). Later, between 1979–1994 inequality increased rapidly. Compared to other countries the Netherlands started from a relatively low inequality, but experienced a relatively high inequality growth (Gottschalk and Smeeding, 2000). Caminada and Goudswaard (2001) found that the two main forces behind this phenomenon are a more unequal distribution of market incomes and changes in social transfers. In 1990 a revision of the tax system led to more inequality. In addition, the growth in the number of two-earner couples increased inequality between 1985 en 1994 (SCP, 2003). Using a decomposition analysis, SCP found that the growth of two-earner couples, in combination with the relatively decreasing incomes of people without employment income, explains about one third of the total increase in inequality. In the second half of the 1990s, income inequality between households decreased slightly (De Vos, 2007), whereas it was quite stable during 2000 to 2007 (Statistics Netherlands).

In the future, the trend of more two-earner households will result in more households receiving two occupational pension incomes. This will increase the average retirement income. With regard to inequality, more two-earner couples can lead to a pooling effect: the inequality within the group of households with two earners is lower than that of households with one earner. This means that an increase in the proportion of two-earner households will, at a certain point, reduce household income inequality.

Data 5.3

The data used in this study are from the 1989–2007 Income Panel Study of the Netherlands (IPO, Inkomens Panel Onderzoek, CBS 2009a) and from the population register (GBA, Gemeentelijke Basis Administratie, CBS 2010). These data were compiled by Statistics Netherlands. Section 5.3.1 provides basic information about IPO and lists descriptives on income and labor market status. Section 5.3.2 describes the information used from the GBA.

Dutch income panel (IPO) 5.3.1

IPO contains information about households and their income, based on administrative data. Most of these data are from the Dutch National Tax Administration. Additional data are derived from the registration of rent subsidies and study grants. In the IPO, so called 'key persons' are randomly drawn from the Dutch population and are followed over time. Data on all household members of the key persons are also available. Major advantages of having administrative data are a very low attrition rate and a high level of representativeness. It is a well-known fact that the rich and the poor are often underrepresented in surveys, institutional households are in general not included, and the elderly population and single person households have relatively low participation rates in surveys (Alessie et al., 1990, Knoef and de Vos, 2009b). Another advantage of administrative data is that the observed variables are measured with a high degree of accuracy. A drawback of the IPO is that it lacks some crucial background variables, such as education levels. Variables that are included in the data are individual characteristics (such as gender, date of birth and marital status), household characteristics (such as family composition) and financial variables related to income.

We have waves from 1989 to 2007 at our disposal, which means that we have data covering 19 years. A revision of IPO in 2000 changed several data sources, definitions and methods. For the year 2000 two datasets are available: one with the data sources, definitions and methods from before the revision, and one with the data sources, definitions and methods after the revision. We have equalized the definitions as much as possible. Appendix A describes the most important changes.

To compare incomes of households with different compositions and size a wide range of equivalence scales is available. Buhmann et al. (2005) review available equivalence scales and find that the choice of the equivalence scale can affect inequality rankings. We chose the equivalence scale proposed by Statistics Netherlands because it is based on the Dutch situation (see Siermann et al., 2004, table 15, for more details). It takes into account the number of adults and the number of children in a household. Kalmijn and Alessie (2008) found that the modified OECD scale and the CBS scale yield very similar results with regard to the distribution of equivalized household income. Appendix B

Table 5.3.1: Data selection[a]

Raw sample	1,835,819
Observations left over after removal (sequentially)	
Household income missing	1,819,048
Age of a household member missing	1,819,007
Negative or zero household income	1,807,963
Households with 9 or more household members	1,802,405
Key persons member of multiple couple household	1,793,807
Key person is a child or a student	1,290,226
Select key persons of age 36–90	958,188
Select key person born between 1917–1970	911,079
Bottom or top 0.1% of income distribution (by year)	909,257
Minus the year 2000 after revision[b]	861,336

[a] The number of key persons and reason of removal from the sample. Key persons are randomly drawn from the Dutch population and are followed over time. We have information about all household members of the key persons.

[b] In the estimations we use the year 2000 before revision, instead of the year 2000 after revision.

explains the definition of income. Income is always deflated/inflated to the prices of 2005 (real income) using the CPI.

Table 5.3.1 reports on the selection of the data. At first, we exclude households with missing or non-positive household income.[2] Furthermore, we exclude households with nine or more household members and households where the key person is a member of a multiple couple household, a child or a student. We select all households where the key person is born between 1917–1970[3]

Table 5.3.2: Descriptives equivalized household income, age key person 50–64[a]

Year	Mean	p10	p50	p90	$\frac{p90}{p10}$	$\frac{p90}{p50}$	$\frac{p50}{p10}$	Gini
1989	20114	11310	18346	30705	2.71	1.67	1.62	0.228
1990	21187	11599	19096	32811	2.83	1.72	1.65	0.241
1991	21220	11464	19139	32566	2.84	1.70	1.67	0.243
1992	21183	11473	19242	32495	2.83	1.69	1.68	0.241
1993	21329	11530	19360	32931	2.86	1.70	1.68	0.241
1994	21241	11200	19210	33107	2.96	1.72	1.72	0.247
1995	21718	11320	19490	34049	3.01	1.75	1.72	0.250
1996	21971	11477	19727	34343	2.99	1.74	1.72	0.251
1997	22073	11530	19943	34418	2.99	1.73	1.73	0.248
1998	22747	12025	20534	35206	2.93	1.71	1.71	0.246
1999	23034	11985	20747	35923	3.00	1.73	1.73	0.253
2000	23596	12297	21190	36589	2.98	1.73	1.72	0.253
2000	23506	12428	21128	35947	2.89	1.70	1.70	0.248
2001	24203	12838	21786	37468	2.92	1.72	1.70	0.247
2002	24407	13024	22077	37580	2.89	1.70	1.70	0.244
2003	24128	12930	21911	37330	2.89	1.70	1.69	0.243
2004	24463	13124	22035	37641	2.87	1.71	1.68	0.245
2005	24589	13102	21994	38118	2.91	1.73	1.68	0.247
2006	23629	12598	20859	36872	2.93	1.77	1.66	0.254
2007	24351	12814	21528	38257	2.99	1.78	1.68	0.258

[a] Source: IPO, own computations. In this study income is always inflated/deflated to 2005 euro's. The year 2000 is presented two times, first for the data before revision and secondly for the data after revision.

[2] In 64% of the households with negative income, there are one or more self-employed household members. However, of all households with self-employment, only 3.3% have a negative income, so that we do not overlook a large part of the self-employed households.

[3] In this way we can make predictions until 2020 for the population of age 50–90, because the cohort born in 1970 reaches the age of 50 in 2020 and the cohort born in 1917 is of age 90 in the last wave of the data (2007). In this study we ignore new immigrant families. For the elderly we expect the effect of this choice to be small.

and is of age 36–90. Finally, households in the bottom or top 0.1% of the income distribution are excluded.

Descriptives household income

Tables 5.3.2 and 5.3.3 describe the distribution of equivalized household income for key persons in the age groups 50–64 and 65–90, respectively. Henceforth, any reference to 'income' should be read as 'equivalized household income'.

During the years 1989–2007, income increased. In the age group 50–64, mean income increased by 21%, from 20,114 euro in 1989 to 24,351 euro in 2007. In the age group 65–90, income was fairly constant during the 1990s. It increased by only 1% between 1990–1999, compared to 9% between 2000–2007.[4]

The Gini coefficient and the decile ratios show that inequality in the age group 50–64 increased between 1989 and 1995 and remained fairly constant thereafter.[5] For the age group 65–90, inequality is lower and shows a different pattern. It grew between 1989–1991, but declined in the years after 1991. Since 1998, inequality in the age group 65–90 has been quite stable. These developments add to the results of Gottschalk and Smeeding (2000), who found that overall income inequality increased between 1979 and the mid-1990s. Several factors may have induced these trends, such as the increased female labor force participation, changes in early retirement schemes, the development of the pension system, and the business cycle.

Labor and occupational pensions

Changes in participation rates influence the income structure of the next generation of pensioners, since more labor income today generally leads to more occupational pension income in the future. Figure 5.1 shows the percentage of females receiving labor income across several generations. For example, '1938'

[4]This is probably related to the fact that no indexation of public pension benefits occurred in the 1990s.

[5]In 1990 a major revision of the tax system took place with distributional consequences ('operatie Oort'). This explains (part) of the difference between inequality measures from 1989 to 1990.

Table 5.3.3: Descriptives equivalized household income, age key person 65–90[a]

Year	Mean	p10	p50	p90	$\frac{p90}{p10}$	$\frac{p90}{p50}$	$\frac{p50}{p10}$	Gini
1989	17031	10355	14699	26732	2.58	1.82	1.42	0.225
1990	17725	10416	14850	28459	2.73	1.92	1.43	0.242
1991	17738	10388	14890	28641	2.76	1.92	1.43	0.244
1992	17626	10542	14935	28176	2.67	1.89	1.42	0.236
1993	17489	10557	14867	27746	2.63	1.87	1.41	0.231
1994	17252	10481	14639	27183	2.59	1.86	1.40	0.231
1995	17278	10605	14659	27246	2.57	1.86	1.38	0.228
1996	17375	10665	14799	27300	2.56	1.84	1.39	0.228
1997	17461	10835	14795	27343	2.52	1.85	1.37	0.225
1998	17916	11275	15192	27758	2.46	1.83	1.35	0.221
1999	17936	11222	15196	27696	2.47	1.82	1.35	0.224
2000	18337	11393	15515	28406	2.49	1.83	1.36	0.228
2000	18541	11504	15708	28337	2.46	1.80	1.37	0.227
2001	18562	11702	15737	28252	2.41	1.80	1.34	0.224
2002	19044	11932	16125	29300	2.46	1.82	1.35	0.225
2003	19065	11956	16180	29193	2.44	1.80	1.35	0.225
2004	19189	12073	16366	29316	2.43	1.79	1.36	0.222
2005	19367	12014	16439	29717	2.47	1.81	1.37	0.228
2006	19575	12247	16771	29788	2.43	1.78	1.37	0.224
2007	20048	12406	17196	30592	2.47	1.78	1.39	0.227

[a] Source: IPO, own computations. In this study income is always inflated/deflated to 2005 euro's. The year 2000 is presented two times, first for the data before revision and secondly for the data after revision.

refers to persons born in 1938. The vertical differences between lines measure the 'cohort-time' effects. We use this terminology to emphasize that it is not possible to disentangle age from cohort and time effects in this figure. Female participation rates have increased considerably and it can be seen that there are important cohort-period effects. This has also been found by Euwals et al. (2011), who claim that changed attitudes towards the combination of paid work and children have played a major role. This trend will have considerable consequences for the income structure of next generations of pensioners, as more two-earner couples today will lead to more couples receiving double pension incomes in the future.

Figure 5.2 shows the percentage of males receiving labor income across several generations. From age 50 to 65 there is a steep decrease in the proportion

Figure 5.1: Percentage of females receiving labor income, per age and cohort

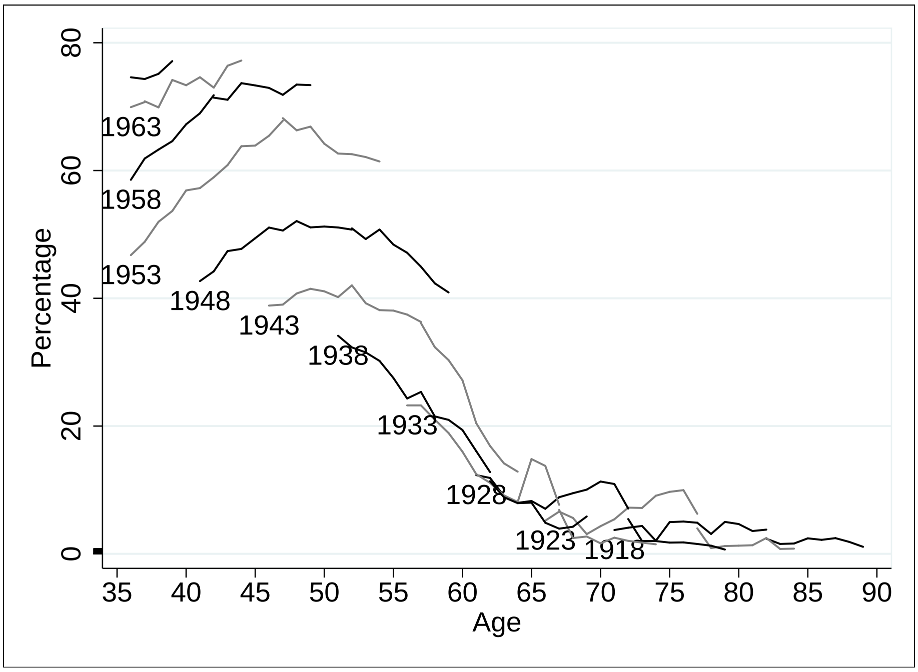
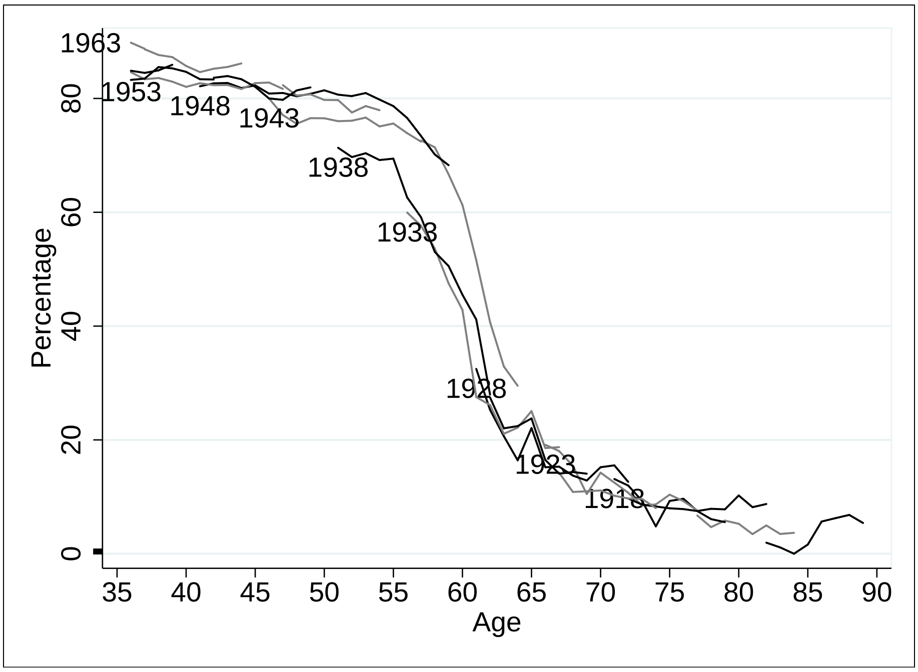


Figure 5.2: Percentage of males receiving labor income, per age and cohort



of males receiving labor income. However, as from the generation born in 1943, more men of age 50–65 continue to participate in the labor market. At age 60, about 46% of the men born in 1938 received labor income, while about 61% of the men born in 1943 received labor income. This means there is a cohort-period effect of 15%-points, probably due to policy reforms that have been implemented since the late 1990s to discourage early retirement. The results are in line with Kapteyn et al. (2010), who found that the labor force participation for men of age 55–64 decreased until 1993 and has increased afterwards. Furthermore, they agree with SCP (2006), who report that between 1994 and 2003 the importance of earnings for the age group 55–64 increased, at the expense of income from early retirement.

Retirement decisions of married men and women are interrelated. We find that the participation rates of married women, younger than 65, who have a husband aged 65 or over have increased considerably, from 12% in 1989 to 23% in 2007. Meanwhile, the age difference between men and women has been quite stable over time. This is an interesting trend, which leads to elderly households receiving more labor income and more occupational pension income in the future.

Coherent with the growth in the female labor force participation, table 5.3.4 shows that especially the percentage of women receiving an occupational pension has increased. In 1989, 29% of the women aged 65 received an occupational pension; in 2007 this was almost 63%. Also with age an increasing percentage of women receive an occupational pension income. For example, in 1989 29% of the 65-year-old women received an occupational pension income. In 1999 these women were 75 year old and the percentage receiving an occupational pension income has increased to 40%. This increase can be attributed to widow pensions, which women start to receive later in life, when their partners die. New generations of retired men also receive occupational pension income more often. The percentage of men receiving an occupational pension at age 65 has increased from 85% in 1989 to 96% in 2007.

Income composition 1989–2007

This study addresses the distribution of household income, which is the sum of several income components. This section investigates the composition of these

Table 5.3.4: Percentage of men and women receiving occupational pension income[a]

Age	Men			Women		
	1989	1999	2007	1989	1999	2007
50	2.00	5.18	3.52	5.89	6.88	5.13
55	8.16	8.20	9.20	10.28	8.19	10.11
60	34.47	33.92	31.45	19.92	18.45	28.68
65	84.76	88.75	96.21	28.94	36.00	62.71
70	84.21	89.69	91.94	41.13	40.83	46.53
75		86.70	90.22		40.48	51.47
80		85.91	87.19		57.03	61.96
85			83.70			69.19
90			88.89			67.47

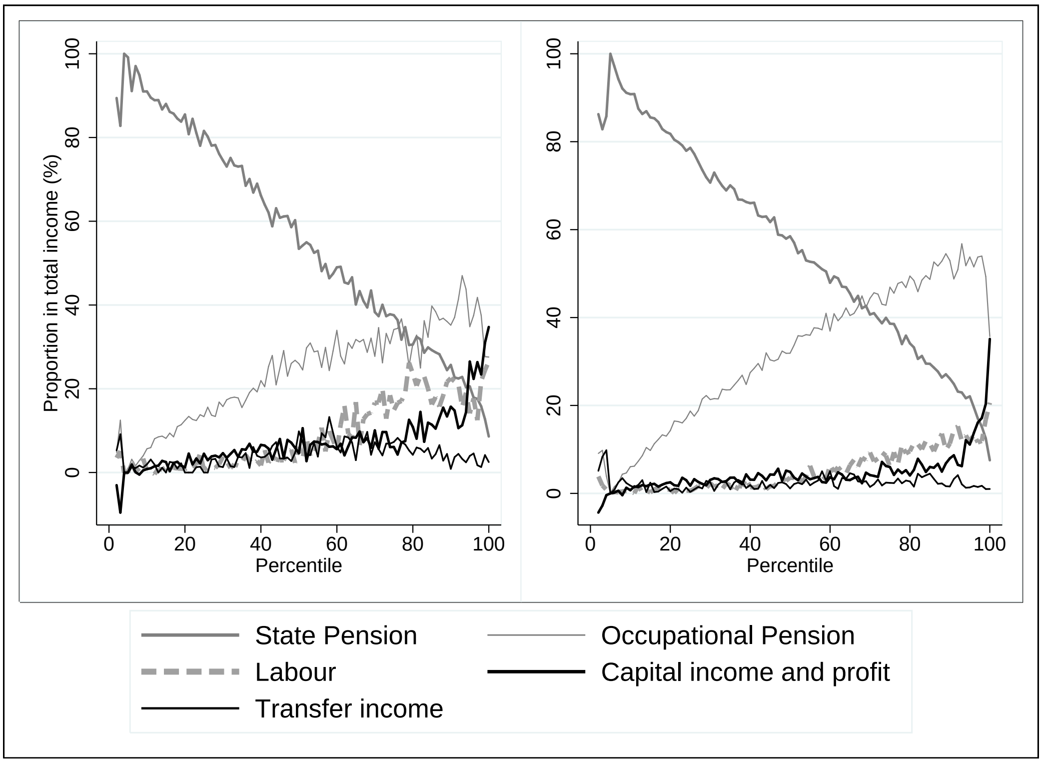
[a] Source: IPO, own computations. The percentage of men and women with occupational pension income in the years 1989, 1999 and 2007. Because we focus on the birth years 1917–1970, there are no descriptives for age 75+ in 1989, and age 85+ in 1999. Widow pensions are included.

income components. Figures 5.3 and 5.4 show the income composition in 1989 and 2007 for the age groups 65–90 and 50–64, respectively. The horizontal axes give the percentiles of the income distribution, the vertical axes the proportions of the various income components.

In figure 5.3 (age 65–90) the income of a median income household consists of 58% public pension benefits, 32% occupational pension income, and 10% remaining income sources. Between 1989 and 2007, occupational pensions became more important for almost all percentiles. In 1989, as from about the 85th percentile occupational pensions were more important than state pensions. In 2007, already as from the 70th percentile occupational pensions are more important than state pensions. This finding might be explained in part by spending cuts of the government during the early 1990s, as a result of which public pension benefits have not been adjusted for inflation.[6] Another explanation for the increase of the income share of occupational pensions is the development of the pension system in the 1950s and the 1960s (Deelen, 1995).

[6]In principle, public pensions follow the gross minimum wages, which are linked to the development of the contractual wages. If no indexation takes place, public pensions will lag behind the growing prosperity; all the more because contractual wages in turn lag behind earned incomes, because of occasional increments and promotions. (De Kam and Nijpels, 1995)

Figure 5.3: Income composition for the age group 65–90 in 1989 (left) and 2007 (right)

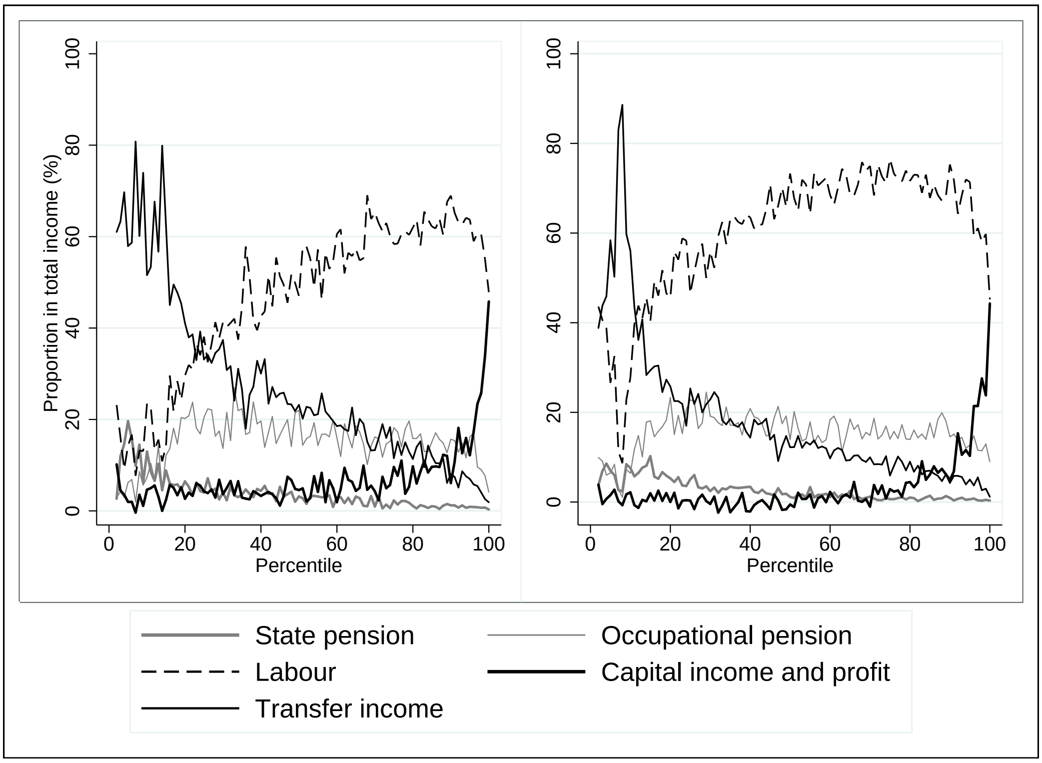


Note: for each percentile in the income distribution these figures show the average proportion of several income components. Transfer income includes welfare, disability benefits, and unemployment benefits.

Above age 65, labor income is most important for households in the high percentiles. It is likely that households with the 'better' jobs remain active on the labor market and obtain a relatively high share of their income from work. Capital income only plays a substantial role for the top 5% of the households.

For age 50–64 we see that the proportion of labor income increases over the income distribution: the more income households receive, the more labor is an important component (figure 5.4). As expected, for the lower percentiles transfer income is important. Interestingly, less households depended on transfer income in 2007 than in 1989. While in 1989, as from the 25th percentile labor income becomes more important than transfer income, in 2007 this intersection already takes place at the 13th percentile.

Figure 5.4: Income composition for the age group 50–64 in 1989 (left) and 2007 (right)



Note: for each percentile in the income distribution these figures show the average proportion of several income components. Transfer income includes welfare, disability benefits, and unemployment benefits. Occupational pensions also contain early retirement income.

5.3.2 Population register (GBA)

Trends in marital status influence the income distribution. Divorces, for example, lead to changes in income as well as household formation (and thus the equivalence scale). The GBA contains information on the marital status of all people registered in Dutch municipalities.[7] Because the GBA contains much more observations than IPO, we use the GBA to estimate transition models of marital status from one year to another. Data is available from January 1 1995 to January 1 2008. Just as in IPO, we select all persons born between 1917–1970 with age 36–90. Furthermore, to be able to estimate transitions between t and $t+1$, the marital status in $t+1$ has to be known. Therefore,

[7]Individuals not registered as residents are, for instance, NATO personnel, diplomats and individuals illegally residing in the Netherlands.

2006 is the last year we can use and persons who, for example, emigrate or decease in $t+1$ are excluded at time t.

We end up with 6,812,340 individuals in 1995, increasing to 8,673,138 individuals in 2006. The percentage of married people who divorce between t and $t+1$ raised from 0.7% in 1995 to 0.8% in 2006. Furthermore, per year on average 2.5% of the divorced persons make a transition into marriage. Most widows and widowers are relatively old and do not remarry again. On average, 0.4% of the widows and widowers make a transition into marriage from one year to the other.

Microsimulation model 5.4

Microsimulation models are used for income predictions and pension issues internationally.[8] These models are in general very demanding multi-year projects which require a lot of data (Harding, 2007). One often needs to combine various data sources with different samples, so that one needs to rely on matching of ‘statistical twins’ (e.g. Geyer and Steiner, 2010) and on surveys that often suffer from representability problems, especially when focusing on the elderly population.[9]

In a microsimulation model the quality of the input data is of prime importance: if the baseline data are not representative, the predictions of the population will not be representative either (Martini and Trivellato, 1997). This study uses a long and representative administrative panel. Although administrative data contain less detailed information on the characteristics of persons and households, the panel aspect of the data allows us to take into account unmeasured variables such as education, ability, and cohort effects.

To simulate the income distribution of the elderly until the year 2020, we use an open dynamic population model with cross-sectional aging. In our model each characteristic for each person is updated each year (dynamic aging). By

[8]Examples are the MIDAS (Microsimulation for the Development of Adequacy and Sustainability) model for Belgium, Italy and Germany (Dekkers et al., 2008). The MINT (Modeling retirement Income in the Near Term) model for the US (Butricia et al., 2001, Panis and Lillard, 1999, Smith et al., 2007, Toder et al., 1999), Pensim2 for the UK (Emmerson et al., 2004), and SESIM which is used to study the income of the Swedish babyboomers (Flood et al., 2008).

[9]Nursing homes are often excluded in surveys and the elderly population living in private households are often underrepresented (Knoef and de Vos, 2009b).

contrast, in microsimulation models with static aging individual characteristics are constant over time. Then, the weights attached to each individual change over time and mimic the process of demographic aging.

Static aging is well suited for short to medium term forecasts (3–5 year), where it can be expected that large changes have not occurred in the underlying population (O'Donoghue, 2001). An example of a model with static aging can be found in Soede et al. (2004), who analyze future incomes in six European countries.

Cross-sectional aging means that we first simulate all individuals for one year, then for the second year, and so fort. Longitudinal simulation models, on the other hand, simulate individual one for all years, the same for individual two, and so forth. Cross-sectional aging allows us to have interactions between household members. For example, husbands and wives make joint labor supply decisions, and the death of a household member can influence the labor market positions of the remaining household members. The model is open, as marriage and birth lead to new synthetic household members. In closed microsimulation models the matching of spouses is restricted to persons within the sample.

Figure 5.5: Design of the microsimulation model

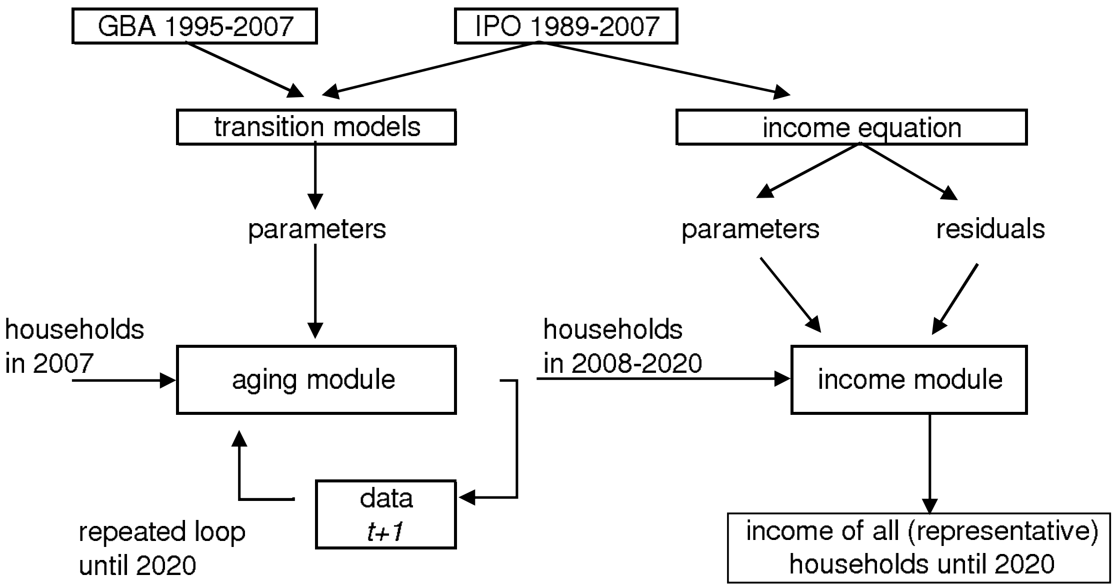


Figure 5.5 describes the design of the model. The representative households in the Dutch Income Panel of the year 2007 are the starting point of the simulation. They form the base population of the model. We dynamically age

all members of these households until 2020 in the aging module. In the aging module, people age, they may decease, divorces may take place, children may leave their parental home, new partners or children may enter the household and labor market positions may change. Transition models, estimated with IPO and GBA data, are used to predict the transitions in household demographics and labor market positions. Furthermore, we take into account differential mortality with regard to income.

After the aging module, households move into the income module, where household incomes are predicted. To this end, we estimate a fixed effects income equation, taking into account age and period effects, household demographics and labor market status. The fixed effects take into account unobserved heterogeneity and we consider the persistency and heteroskedasticity of income shocks.

In the remainder of this section we explain the income equation (5.4.1), the implementation of differential mortality in the aging module (5.4.2), the transition models with regard to household demographics (5.4.3) and the transition models with regard to labor market status (5.4.4).

Income equation 5.4.1

To predict income trends for future generations of pensioners we model household income using a fixed effects model. We include age effects, period effects, and socio-economic variables in our regression model. Socio-economic variables enable us to take into account developments in the income distribution due to different socio-economic characteristics of future pensioners. The fixed effects allow us to control for time-invariant omitted variables that influence the income of a household. They include education, ability, and income differences between cohorts caused by productivity differences.[10] Fixed effects are in line with Haveman et al. (2007), who found that preretirement economic advantages continue into retirement. We have found only one other microsimulation model using fixed effects. That is the MINT model that uses fixed effects to take into account unmeasured heterogeneity in lifetime preretirement earnings profiles (Butricia et al., 2001, Toder et al., 1999).

[10]Kapteyn et al. (2005) found that productivity growth can explain all generation effects with regard to income.

We prefer a fixed effects model to a random effects model as household specific effects (μ_i) may be correlated with the included covariates. For example, 'ability' is likely to be correlated with labor market status. The disadvantage of a fixed effects estimator in microsimulation models is that it rules out out-of-sample simulations (Wolf, 2001). However, in this analysis we can use a fixed effects model because our target population are future pensioners, who are already born and available in the data.

Income profiles are estimated with the same data as the base population is derived from. The fixed effects income equation is

$$y_{it} = \alpha + \beta' x_{it} + \mu_i + \nu_{it}, \tag{5.1}$$

where y_{it} is the 'log' of equivalized household income[11] of household i in time period t, α is a scalar, x_{it} is the it-th observation on K explanatory variables, β is a parameter vector of size K, μ_i is the unobserved individual effect and ν_{it} is the error term. We assume strict exogeneity

$$E(\nu_{it} \mid \mu_i, x_{i1}, ..., , x_{it}, ..., x_{iT}) = 0 \tag{5.2}$$

and identify α using the normalization $\sum_{i=1}^{N} \mu_i = 0$. The estimation of α, β and μ_i is explained in Appendix 5.C. We estimate three specifications of the income equation. In the first specification, the vector x_{it} only contains age and period effects. This is the pure specification where income mobility only results from income shocks. In the second specification, demographic variables such as household size and marital status are added, and in the third specification also the labor market positions of household members are taken into account. By adding extra variables to the vector x_{it}, more individual heterogeneity is introduced in the income path. Adding household size as an explanatory variable, in addition to the use of the equivalence scale in the dependent variable, leads to information about the income effect of an extra men, women or child in the household. For example, if the coefficient for the number of adult men in the household is positive, we can conclude that on average, the income of an extra men exceeds his marginal costs of living (determined by

[11] If one wants to simulate income components separately, one should take into account the correlations between components. There is no need for that in this study.

the equivalence scale).

Age and period effects are implemented as dummy variables, so that their relationship with income is very flexible. However, age, period, and cohort effects (cohort effects are captured in the individual effect) cannot be identified empirically, since calendar time is equal to the year of birth plus age. We follow the identification restriction proposed by Deaton and Paxson (1994), which means that we assume that all time dummy coefficients add up to zero and are orthogonal to a linear time trend. We assume that all period effects are due to unanticipated business cycle shocks.

Households experience income shocks, the size of which may depend on characteristics of the household (heteroskedasticity). For example, income shocks may be larger during working life than during retirement, and may be higher for singles than for couples. Furthermore, the question arises how long income shocks persist (autocorrelation), and whether the persistency of a shock depends on the position in the lifecycle.

When a household experiences an income shock in period t, this may have an effect on the income in the periods following t. The error term v_{it} therefore might follow an autoregressive scheme. To model this we fit the following auxiliary regression model of order two[12]

$$v_{it} = \rho_{1,it} v_{i,t-1} + \rho_2 v_{i,t-2} + \varepsilon_{it}, \tag{5.3}$$

where we assume ε_{it} to be serially uncorrelated. The persistency of a shock may depend on the position in the lifecycle.[13] Therefore, we allow $\rho_{1,it}$ to be a function of age.[14]

$$\rho_{1,it} = \rho_{0,1} + \rho_{1,1} \frac{\text{age}_{it}}{10} + \rho_{2,1} \left(\frac{\text{age}_{it}}{10} \right)^2 \tag{5.4}$$

As explained above, the variance of an income shock may depend on the characteristics of a household. We take this heteroskedasticity into account by investigating the distribution of ε_{it} for several mutually exclusive groups

[12]We find that higher orders are of no importance.

[13]Kalmijn and Alessie (2008) found that the two-year autocorrelation of equivalized income is quite stable during midlife, but moves to a higher level after age 65.

[14]We have tried several specifications and have also investigated whether it is relevant to specify ρ_2 as a function of age.

of households. For example, the group of households where the key person is younger than 65 and the group of households where the key person is older than 65. For each group we draw income shocks from the empirical distribution of residuals in 2001–2007 for that group ($\widehat{\varepsilon}_{i,2001},\ldots,\widehat{\varepsilon}_{i,2007}$).[15]

In the predictions we assume period effects to be zero, such that the predicted incomes are free from the effects of the business cycle. Finally, we take into account that as from 2015, the partner bonus for the younger partners of state pension beneficiaries with no or low income will be abolished. We subtract the partner bonus for all households who are not eligible for a partner bonus anymore, and of whom the younger member of the couple has no labor income. SZW (2009) found that remaining household income for most of these households will not reach the eligibility limit for social assistance.

5.4.2 Differential mortality

In the aging module, where we age all household members in the microsimulation model from 2007 to 2020, persons may decease. To determine whether an individual in the sample deceases we apply Monte carlo simulations (see e.g. Law and Kelton, 1982). For each individual j and each period t from 2008 to 2020 we draw a random value m_{jt} from the uniform distribution. If m_{jt} is lower than the predicted mortality rate the individual deceases.

We use predicted mortality rates per age, cohort, and gender published by Statistics Netherlands and adjust the mortality rates of the first and fourth income quartile using the degree of differential mortality found by Kalwij et al. (2009) for the Netherlands. If we would not take into account differential mortality we would underestimate the income level of the elderly, as low income households would survive relatively too often and high income households would survive not often enough.[16] Kalwij et al. (2009) find a quartile ratio Q1/Q4 of 2.2 for men and 1.7 for women as from age 65, meaning that mortality rates in the first income quartile are 2.2 times higher for men and 1.7 times higher for women, relative to the fourth quartile.[17] As from age 65 we

[15] By using the years as from 2001, possible effects caused by the revision of the data are excluded.

[16] When we do not take into account differential mortality, the average yearly income growth of pensioners between 2008 and 2020 is 0.15%-points lower.

[17] This is in line with findings in other European countries, e.g. Von Gaudecker and Scholz (2007) for Germany and Osler et al. (2002) for Denmark.

therefore adjust mortality rates such that mortality rates in the first quartile are 2.2 (or 1.7 for women) times higher than in the fourth quartile, keeping the average mortality rate equal. Before age 65 mortality rates are small, such that differential mortality will not make a relevant difference.

To determine which households belong to the first and the fourth income quartile, we use the fixed effects estimation of the first specification (where only age and period effects are taken into account). The fixed effects of this estimation give us a measure of the ‘lifetime income position’ of households, as in this specification a correction has been made only for the age profile, the business cycle, and for income shocks.

Transitions in marital status and household composition 5.4.3

Using GBA data, we model the following transitions in marital status from year to year: married-divorced, unmarried-married, widow(er)-married, and divorced-married. Logit models are employed to estimate the transition probabilities between the various marital statuses. The transition models are estimated for men and women separately and use age and year of birth as explanatory variables. We do not explicitly model transitions into widowhood. Becoming a widow(er) depends on the death of a partner. This probability is incorporated via mortality (described in section 5.4.2).

We assume people to make at most one transition in marital status per year and apply Monte Carlo simulation to assess whether a change indeed occurs. In case of a divorce, the partner of the key person is removed from the household, and in case of marriage a new household member is added. These new household members have the same age as their partners and the opposite gender.

The probability of a child leaving the parental home from one year to the other is estimated using a logit model, where age and gender of the child are the explanatory variables. The probability of a newborn child is also modeled with a logit model. The explanatory variables are the age and gender of the key person in the household, whether there is a couple in the household, and the number of children which are already present in the household. In reality, also older children may enter the household. The simulation model ignores children already born to enter the household.

5.4.4 Transitions in labor market status

The third specification of the income equation distinguishes three labor market positions: (1) receiving labor income, (2) receiving occupational pension income and (3) receiving none of these two ('other'). In order to belong to (1) or (2), labor income or occupational pension income has to be at least 500 euro per year. In case an individual receives both labor income and pension income the highest income component counts.

We model the transitions between the three labor market positions. Herewith, we assume 'occupational pension' to be an absorbing state. The labor market positions of the two members of a couple are interrelated. We therefore estimate transition models for singles and couples separately. Concerning singles, we estimate multinomial logit models for men and women one by one. For couples we treat the three labor market outcomes of a husband and a wife as $3 \times 3 = 9$ univariate outcomes. For instance, we model the transition probability from the state where both husband and wife work to the state where only the wife works and the husband receives an occupational pension income. Transitions between the nine states are estimated with multinomial logit models. The explanatory variables used in the estimations are age, cohort, marital status, and the number of children. Using the parameters of the transition models, we estimate the transition probabilities for all singles and couples, given their age, marital status, and labor market position in the previous period. Also here, random draws from the uniform distribution determine whether a transition takes place.

To determine the labor market status at time $t + 1$ with the transition models, we need the labor market status at time t. A problem arises for new household members and children who enter adulthood. To determine an initial state for them we estimate a multinomial logit model per gender, with age and cohort as explanatory variables. The increased labor market participation of women therefore enters the model in two ways: via the initial labor market positions of women and via cohort effects in the labor market transition models.

5.5 Estimation results

This section discusses the estimation results of the income equation explained in section 5.4.1. The estimation results of the transition models described in section 5.4.3 and 5.4.4 are discussed in Appendix 5.D.

Table 5.5.1 shows the estimation results of the fixed effects income equation. The first two columns show the estimation results of the first specification, where only age and period effects are taken into account. In the second specification, household demographics are added and in the third specification also labor market positions are added. All coefficients are based on within variation.

Table 5.5.1: Estimation results fixed effects income equation[a]

	Coef. 1	S.e.	Coef. 2	S.e.	Coef. 3	S.e.
age dummies[b]	yes		yes		yes	
year dummies	yes		yes		yes	
# adult men			0.131	0.0018	0.037	0.0028
# adult women			0.061	0.0018	-0.030	0.0023
# children			-0.068	0.0015	-0.059	0.0015
widower			0.138	0.0072	0.084	0.0071
widow			0.044	0.0051	-0.044	0.0058
divorced (man)			0.033	0.0062	0.021	0.0060
divorced (woman)			-0.123	0.0077	-0.140	0.0076
unmarried (man)			0.057	0.0091	0.050	0.0088
unmarried (woman)			-0.071	0.0120	-0.080	0.0119
# labor (man)					0.120	0.0026
# labor (woman)					0.118	0.0018
# occ. pension (man)					0.058	0.0032
# occ. pension (woman)					0.099	0.0034
$\rho_{0,1}$	-0.160	0.0253	-0.217	0.0259	-0.227	0.0260
$\rho_{1,1}$	0.162	0.0091	0.163	0.0093	0.162	0.0093
$\rho_{2,1}$	-0.010	0.0008	-0.009	0.0008	-0.009	0.0008
ρ_2	0.066	0.0012	0.054	0.0012	0.055	0.0012
α	9.909		9.746		9.805	
σ_μ	0.370		0.369		0.342	
σ_ε	0.210		0.205		0.234	
R^2	0.061		0.134		0.165	
N	861336		861336		861336	

[a] Reference categories are ‘age 65’ and ‘married’. For the identification of age, period, and cohort effects the method of Deaton and Paxson (1994) is used. Clustered standard errors are used to take into account the correlation of the error terms in the same household.
[b] The coefficients of the age specific dummy variables can be found in Appendix 5.E

In all three specifications, age effects increase until about age 55 and decrease afterwards. As from age 70 they increase again, even when we control for selectivity by adding selection dummies. The shape of the age profiles of specification two and three are very similar, while the age profile of the first specification is more pronounced. The first specification has relatively high age effects around age 54, caused by children leaving their parental home. Equivalized household income increases when children leave their parental home, as the equivalence scale captures the fact that children cost money.[18] Specification two and three correct for the presence of children, hence they have lower age effects around age 54. The estimated period effects follow the development of the business cycle.

Specification two shows that households with more adults have on average a higher equivalized household income. On average adults thus yield more income than 'costs' (in terms of the increase in the equivalence scale). Households with more children, on the other hand, have a lower equivalized income. Kalmijn and Alessie (2008) found that this is partly due to a decline in the personal income of women after the birth of children, but mainly because of the increase in expenditures compelled by having children.

Marital status is significantly associated with income. Compared to divorced men, divorced women are relatively worse off. A divorce often coincides with a loss of an adult in the household, such that the total effect of a divorce for men is a 2.8% loss of income (0.033-0.061) and for women a 25% loss of income (-0.123-0.131). Widowers and widows are better off than unmarried men and women, and the unmarried are on average better off than divorced men and women. Compared to marriage, men have on average 8% more income in widowhood, but women are married 9% financially better off than in widowhood.

The final specification takes labor market positions into account, namely, the number of men and women receiving labor income, and the number of men and women receiving occupational pension income. These variables can be considered endogenous explanatory variables, but although this implies the model estimates cannot be given a causal interpretation, they can be used for predicting income. As expected, the number of working men and women increases household income. The number of men and women with an occupational pension increases household income in a smaller degree.

The parameters $\rho_{0,1}$ to ρ_2 in table 5.5.1 show that income shocks are persistent and that persistency increases with age. In the first specification ρ_1 (the first-order coefficient defined by equation (5.4)) increases from 0.29 at

[18] Money transfers between parents and children not living in the same household cannot be taken into account because they are not available in the data.

age 36 to 0.50 at age 80. In the second and third specification $\rho_{0,1}$ is smaller, such that ρ_1 is somewhat smaller until age 70. This can be explained by the fact that the added demographic variables and labor market status capture part of the persistency. Consider for instance a person faced with a negative income shock from a transition to unemployment. Specification three takes labor market status into account, so as long as the person stays unemployed the negative income effect persists. In the first two specifications labor market positions are not taken into account explicitly. However, a person can receive a negative income shock, which may implicitly be caused by unemployment. The parameters $\rho_{0,1}$ to ρ_2 determine the persistency of the shock. This persistency increases with age, comparable to the duration of unemployment, which also tends to increase with age. Finally, σ_μ and σ_ε show that the individual variation is larger than the random component.

Future income shocks are drawn from the empirical distribution of the idiosyncratic residuals in 2001–2007. As shown by Kalmijn and Alessie (2008), the variance of equivalized income (logged) is relatively low after 65. We therefore distinguish between households with key persons younger and older than 65. The standard deviation of the residuals is 40% higher for households where the key person is younger than 65. In the third specification we also distinguish households that do receive labor or occupational pension income from those that do not receive any of these income components. For households where the key person is younger than 65, the standard deviation of the residual is 49% higher in households without labor or occupational pension income, compared to households with labor or occupational pension income. In households where the key person is older than 65, the standard deviation of the residual is 71% higher for households without occupational pension income, compared to households with an occupational pension. In the simulation these results lead to higher income shocks for young households and for households without labor and/or occupational pension income.

Simulation results 5.6

Corresponding to the three specifications of the income equation, we have three predictions of the income distribution until 2020. Before explaining the income predictions we describe the predictions of marital status and labor market status from the aging module, as they are input for the income predictions in the income module. Predictions of marital status are given in table 5.6.1 for the age groups 50–64 and 65–90.

Table 5.6.1: Predictions of marital status[a]

	Men				Women			
Year	Marr	Unmarr	Wid	Div	Marr	Unmarr	Wid	Div
Age 50–64								
2008	75.6	9.9	2.0	12.5	71.8	7.5	6.2	14.5
2009	74.7	10.5	2.0	12.8	70.9	7.9	6.1	15.1
2010	73.4	11.5	2.0	13.0	70.3	8.4	5.9	15.3
2011	72.2	12.1	2.0	13.7	69.5	9.0	5.7	15.7
2012	71.1	12.9	2.0	14.0	68.9	9.6	5.4	16.1
2013	70.1	13.6	1.9	14.4	68.0	10.1	5.5	16.4
2014	69.1	14.4	1.8	14.7	67.1	10.7	5.2	17.0
2015	67.7	15.4	1.8	15.1	66.3	11.3	5.0	17.4
2016	66.4	16.3	1.9	15.5	65.2	12.1	5.0	17.7
2017	65.4	17.0	1.9	15.8	64.4	12.7	4.9	18.0
2018	64.3	17.8	1.9	15.9	63.7	13.3	4.8	18.3
2019	63.1	18.8	1.7	16.4	62.8	14.2	4.6	18.4
2020	61.8	19.9	1.7	16.7	61.8	15.1	4.4	18.7
Age 65–90								
2008	74.5	5.6	12.4	7.5	46.4	5.7	39.5	8.4
2009	73.8	5.7	12.6	7.9	46.9	5.5	39.0	8.6
2010	73.7	5.7	12.4	8.2	47.9	5.5	37.7	8.9
2011	73.4	5.8	12.2	8.7	48.2	5.5	36.9	9.4
2012	73.0	5.8	12.1	9.2	49.1	5.3	35.7	9.8
2013	72.4	5.9	12.1	9.6	49.8	5.4	34.7	10.2
2014	71.9	5.9	12.2	10.0	50.4	5.3	33.8	10.5
2015	71.7	6.0	12.0	10.2	50.4	5.3	33.3	11.0
2016	71.0	6.3	12.2	10.5	50.5	5.3	32.8	11.4
2017	70.9	6.4	12.1	10.6	50.6	5.5	32.2	11.7
2018	70.3	6.6	12.0	11.1	50.7	5.6	31.5	12.2
2019	69.1	7.1	12.4	11.4	50.5	5.8	31.1	12.7
2020	68.4	7.4	12.4	11.8	50.2	5.9	30.8	13.2

[a] Marital status for men and women. E.g. in 2020 about 61.8% of the men in the age group 50–64 will be married.

In the age group 50–64, the most important finding is the growth in the share of divorced people. The share of divorced men increases from 13 to 17%, while that of women increases from 15 to 19%. In the age group 65–90 we find that widowhood among women decreases. This has to do with the converging life expectancies of men and women, which leads to younger cohorts of women being widowed less often. Furthermore, the fall in widowhood can be attributed to the babyboom generation reaching age 65. Therefore, the total age group 65–90 starts to contain relatively many 'young' elderly who are widowed less often (composition effect).

Table 5.6.2: Predictions of labor market status[a]

	Men			Women		
Year	Labor	Occupational pension	Other	Labor	Occupational pension	Other
Age 50–64						
2008	62.6	19.6	17.8	46.3	15.0	38.6
2009	62.8	21.2	16.1	47.4	16.6	36.0
2010	62.3	23.0	14.7	48.4	17.9	33.7
2011	63.1	23.3	13.7	49.9	19.0	31.1
2012	63.3	23.6	13.1	51.9	19.8	28.2
2013	64.4	23.7	11.9	53.6	20.6	25.8
2014	64.5	24.2	11.3	55.1	21.4	23.5
2015	64.6	24.6	10.8	56.1	22.4	21.6
2016	65.0	25.0	10.1	56.7	23.5	19.8
2017	65.3	25.3	9.5	57.4	24.3	18.3
2018	65.8	25.0	9.2	58.1	25.1	16.7
2019	66.2	25.1	8.7	59.1	25.7	15.2
2020	66.3	25.4	8.4	59.5	26.7	13.9
Age 65–90						
2008	3.6	87.0	9.4	2.1	54.0	43.8
2009	3.2	87.5	9.3	2.0	54.8	43.1
2010	3.1	88.0	8.9	2.2	55.3	42.5
2011	3.6	88.0	8.4	2.6	56.1	41.3
2012	4.3	87.6	8.1	3.0	56.9	40.2
2013	4.2	88.1	7.7	3.2	58.5	38.3
2014	4.1	88.8	7.1	3.1	60.1	36.8
2015	4.4	88.9	6.6	2.9	62.0	35.1
2016	4.3	89.5	6.2	2.9	63.9	33.2
2017	4.1	89.8	6.1	3.0	65.4	31.6
2018	4.3	90.0	5.6	3.0	67.3	29.7
2019	4.1	90.6	5.3	3.2	68.6	28.2
2020	4.0	90.9	5.1	3.1	70.4	26.5

[a] In case a person receives both labor income and occupational pension income the labor market status is based on the highest income component. E.g. in 2020 labor is the most important income source for 66.3% of the men in the age group 50–64.

Table 5.6.2 presents predictions of labor market status. For both men and women, and both age groups 50–64 and 65–90, the share of people receiving occupational pension income increases. This especially holds for women, as a result of their strong increase in the labor force participation.

Using the predictions of marital status and labor market status described above, we predict equivalized household income for all households. Table 5.6.3 shows the results of the most extensive prediction, where household demographics and labor market positions are taken into account (model specification three). Incomes in these tables are free from period effects, such as the effects of the business cycle.

According to the predictions, income will increase on average by about 0.6% per year for the age group 50–64 and 1.0% per year for the age group 65–90 between 2008 and 2020. The Gini coefficient and the decile ratio p90/p10 show that inequality in the age group 65–90 will increase until about 2012, to stabilize thereafter. Focussing on the decile ratios p90/p50 and p50/p10, two contradictory developments seem to occur: an increasing inequality in the lower part of the income distribution and a decreasing inequality in the upper part of the income distribution. This shows the importance of investigating the entire income distribution by microsimulation, rather than just investigating the development of an inequality measure such as the Gini coefficient. Inequality indices differ in their sensitivities to income differences in different parts of the distribution, but one index cannot show the different developments occurring throughout the entire income distribution. For the age group 50–64 the Gini coefficient and the decile ratio p90/p10 show that inequality decreases until 2012, but increases afterwards. After 2012 inequality rises in the upper part of the distribution as well as in the lower part of the distribution.

Figure 5.6 shows realizations and predictions of log income per age and cohort. For every cohort the figure presents the income of the 10th, the median, and the 90th percentile. Period effects are excluded for the predictions (the dashed lines) as well as for the realizations (the solid lines). We use *log* equivalized household income, as it is more interesting to compare relative than absolute changes. The age profile of the median incomes and the 90th percentile is stronger than that of the 10th percentile. As expected, younger cohorts have higher incomes than older cohorts. However, for the 10th percentile cohort-time effects decrease between 2008 and 2020, while they do not

Table 5.6.3: Predictions of equivalized household income[a]

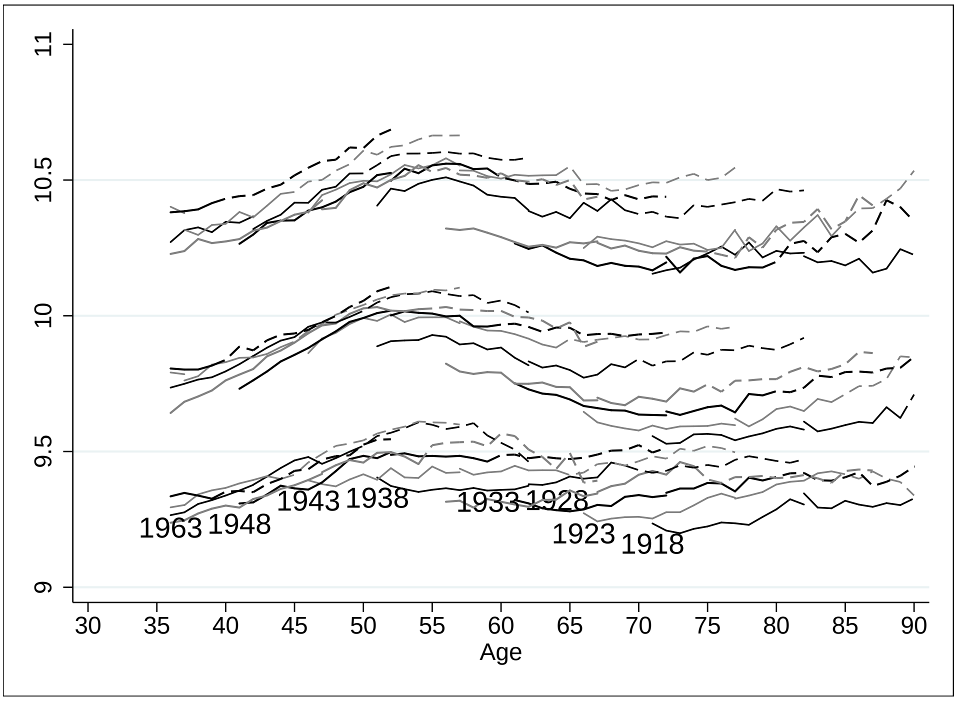
Year	Mean	p10	p50	p90	$\frac{p90}{p10}$	$\frac{p90}{p50}$	$\frac{p50}{p10}$	Gini
Age 50–64								
2008	24484	13282	22028	37983	2.86	1.72	1.66	0.242
2009	24665	13473	22285	38117	2.83	1.71	1.65	0.237
2010	24789	13573	22443	38283	2.82	1.71	1.65	0.237
2011	24877	13757	22594	38348	2.79	1.70	1.64	0.235
2012	25054	13787	22760	38469	2.79	1.69	1.65	0.235
2013	25309	13986	22839	38960	2.79	1.71	1.63	0.238
2014	25506	14015	22961	39209	2.80	1.71	1.64	0.239
2015	25592	14062	23047	39493	2.81	1.71	1.64	0.239
2016	25672	13936	23004	39560	2.84	1.72	1.65	0.243
2017	25719	13937	23000	39649	2.84	1.72	1.65	0.243
2018	25953	13880	23264	40443	2.91	1.74	1.68	0.247
2019	26059	13929	23400	40473	2.91	1.73	1.68	0.248
2020	26239	13946	23387	41488	2.97	1.77	1.68	0.251
Age 65–90								
2008	20267	12214	17805	31156	2.55	1.75	1.46	0.225
2009	20611	12147	18122	31940	2.63	1.76	1.49	0.230
2010	20875	12280	18443	32162	2.62	1.74	1.50	0.229
2011	21252	12377	18862	32715	2.64	1.73	1.52	0.229
2012	21508	12437	19188	33144	2.66	1.73	1.54	0.229
2013	21754	12522	19396	33332	2.66	1.72	1.55	0.229
2014	21951	12734	19612	33716	2.65	1.72	1.54	0.227
2015	22212	12829	19890	34059	2.65	1.71	1.55	0.227
2016	22301	12860	19943	34105	2.65	1.71	1.55	0.228
2017	22436	12836	20111	34332	2.67	1.71	1.57	0.230
2018	22589	12944	20331	34457	2.66	1.69	1.57	0.229
2019	22672	12949	20434	34463	2.66	1.69	1.58	0.229
2020	22874	12976	20632	34925	2.69	1.69	1.59	0.230

[a] In this study income is always inflated/deflated to 2005 euro's. This table shows the results of the most extended model specification, where demographic variables and labor market positions are taken into account (model specification three). The results of the first and second specification can be found in Appendix 5.F.

decrease for the median income households. It becomes clear that the income growth is not the same for everyone.

To show this more thoroughly, figure 5.7 presents the growth of the 10th, median, and 90th percentile of the income distribution between 1989–2020 for the elderly of age 65–90. Like in the other figures, period effects are excluded for the predictions and the realizations.

Figure 5.6: Log equivalized household income per age and cohort.

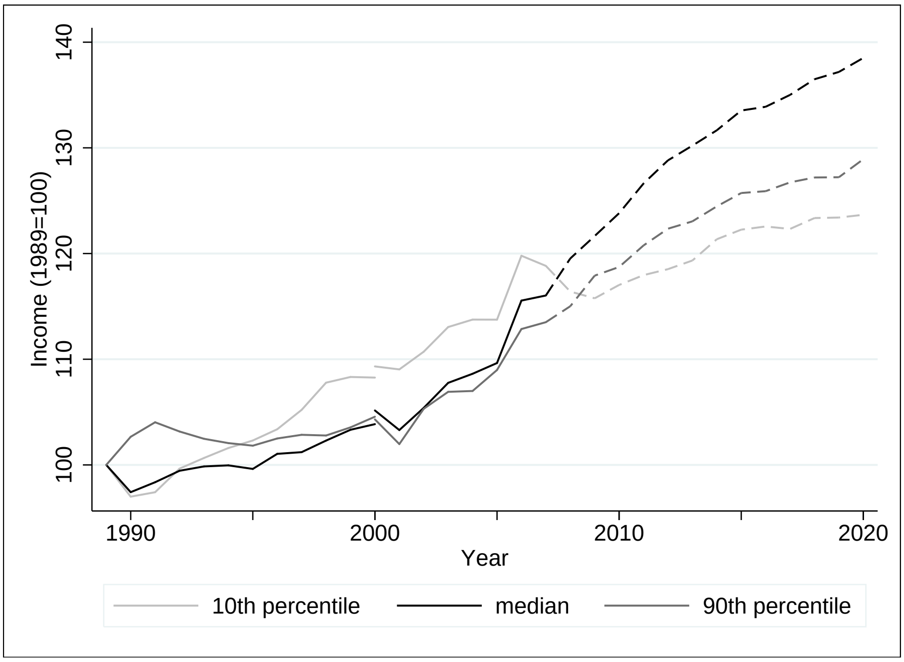


Note: the 10th, median, and 90th percentile of log equivalized household income per age and cohort. The solid lines are realizations corrected for period effects, the dashed lines are predictions made with the most extended version of the microsimulation model (specification three).

Pensioners with median household income experience the highest income growth. As a result, inequality (indeed) increases in the lower part of the distribution and decreases in the upper part of the distribution. Relative poverty thus increases. On average the income growth of pensioners will be higher in the future than it was in the past. When we compare the realized average income growth of pensioners between 1989–2007 with the predicted average income growth between 2007–2020 in figure 5.7, we find an increase in the average income growth per year for median income households from 0.8% until 2007 to 1.4% after 2007. The average income growth of the 90th percentile also increases, from 0.7% per year until 2007 to 1.0% after 2007. The 10th percentile experiences a decrease in the average growth rate from 0.9% to 2007 to 0.3% after 2007. The results of specification one and two are presented in Appendix 5.G and lead to similar conclusions.[19]

[19]The first and the second specification lead to rather similar conclusions, indicating that the use of fixed effects and modeling the error terms make the explicit modeling of demographic

Figure 5.7: Indexed growth of equivalized household income for the elderly of age 65–90



Note: income growth for the 10th percentile, the median and the 90th percentile. The solid lines are realizations corrected for period effects, the dashed lines are predictions made with the most extended version of the microsimulation model (the third specification).

The lower part of the income distribution experiences a relatively low income growth. In this part of the distribution there are many households without occupational pension income. The question arises whether the growing inequality in the lower part of the distribution is caused by an increase in the inequality between households with and without occupational pension income. To answer this question we do a Theil decomposition, concentrating on the lower half of the income distribution. Appendix 5.H describes the Theil decomposition method and table 5.6.4 shows the results.

In the lower half of the income distribution, 21% of the households receive no occupational pension in 2010. In 2020 this proportion will shrink to about 15%. As expected, average income is higher for households with occupational pension income, compared to the households without occupational pension income. The Theil index is about two times higher for households without occupational pension income, but the inequality growth between 2010 and

changes not very important for investigating the future income distribution.

2020 is higher for the households with occupational pension income. The Theil decomposition shows that in 2010, 11% of the inequality in the lower half of the distribution is caused by the inequality between the group of households with and without occupational pension income. By 2020 this is reduced to 5%. The increased inequality in the lower part of the distribution is thus not caused by a higher inequality between households with and without occupational pension income. Instead, the inequality between these two groups will decrease. This means that inequality between households with occupational pension income on the one hand and inactive/self-employed households without pension arrangements on the other will not increase.

Table 5.6.4: Theil decomposition of equivalized household income[a]

Year	2010	2015	2020
% Households without occupational pension	21	18	15
Average income, households without occ. pension	12608	13448	13859
Average income, households with occ. pension	14825	15776	16030
Theil index, households without occ. pension	0.033	0.039	0.039
Theil index, households with occ. pension	0.013	0.016	0.022
Within group inequality	0.0167	0.0197	0.0240
Between group inequality	0.0020	0.0017	0.0012
% Between group inequality	11	8	5

[a] This table concentrates on the lower half of the income distribution of pensioners (age 65–90). It shows the inequality within and between households with and without occupational pension income.

5.7 Conclusions

This chapter examines the income distribution of the Dutch elderly between 1989–2007 and predicts the income distribution of the elderly between 2008–2020. In the predictions we take into account developments in household compositions, developments in labor market positions, productivity differences between cohorts resulting in income differences, differential mortality, and increased longevity.

Predictions are made using an open dynamic microsimulation model with cross-sectional aging. Methodologically this microsimulation model deviates from the more traditional dynamic microsimulation models, by using fixed effects to take into account unmeasured heterogeneity in total household income and taking into account the persistency and heteroskedasticity of income shocks. Using these techniques, one needs less information on all underlying processes influencing income such that administrative data sources, instead of more detailed surveys, can be used. Administrative data are often more representative than surveys, which is important to make reliable predictions for the whole population. We find that income shocks are persistent and that persistency increases over the lifecycle (even after the correction for fixed effects). The variance of income shocks is larger for working-age households than for retirement-age households, and is relatively large for households without labor and/or occupational pension income.

Exploration of the data reveals that between 1989–2007, equivalized household income of the elderly in the age group 50–64 increased on average by 1.1% per year. Income inequality in this age group increased between 1989–1995 and was fairly stable afterwards. The income of the elderly in the age group 65–90 remained fairly constant during the nineties and grew on average by 1.3% per year between 2000–2007. Income inequality in this age group increased between 1989–1991, but decreased thereafter. Occupational pensions became more relevant for the whole income distribution between 1989 and 2007.

The results of the microsimulation model indicate that average income increases for future generations of pensioners. More specifically, we find that between 2008 and 2020 household income increases on average by about 0.6% per year in the age group 50–64 and 1.0% for pensioners of age 65–90. Income growth is not the same for everyone. Among pensioners of age 65–90, households with median income experience the highest income growth. During the years 2008–2020 their income is predicted to grow on average by 1.2% per year, while this is 1.0% for the 90th percentile and only 0.5% for the 10th percentile.

Inequality indices such as the decile ratio p90/p10 and the Gini coefficient show that inequality among pensioners of age 65–90 increases up to 2012 and stabilizes thereafter. However, a closer inspection of the whole distribution reveals that inequality grows in the lower part of the distribution, while it declines in the upper part of the distribution. The growing inequality in the lower half of the income distribution is not caused by an increasing inequality between households with and without occupational pension income. Instead, inequality between households with and without occupational pension income will decrease. The contradictory movements in the lower and upper part of the distribution underline the importance of investigating the whole income distribution, here achieved by using microsimulation, instead of just analyzing the development of one inequality index such as the Gini coefficient. Income mobility is an aspect for future research.

In the introduction we described two examples of proposed policy reforms to keep the costs of the aging society affordable. The results suggest that a policy such as 'fiscalization' can be effective, as a majority of the future pensioners will be considerably wealthier than the current ones. Obviously, this policy measure will not further increase the income inequality among the elderly. Lowering the indexation ambition of the (flat rate) public pensions especially affects the households on the lower end of the distribution and increases inequality further. However, when occupational pensions are also lowering their indexation ambitions (plausible as a result of the financial crisis), inequality remains undisturbed.

5.A IPO before and after the revision in the year 2000

In this appendix we list important changes which took place during the revision of IPO. In addition we mention several steps we have taken to make the data before and after revision more comparable.

- As from 2000 also one-off income such as severance pays are included. (All income that previously belonged to 'bijzonder tarief', it concerns large amounts that do not occur frequently.)
- As from 2000 new data sources are in use. In particular with regard to rents and dividends. Before 2000 we did not observe rents and dividend for the people who were only obliged to pay tax on wages (rents and dividends belonging to the tax free allowance). As from 2000 new data sources are used such that rents and dividends are observed for everyone (small amounts of income for a rather large group of observations). In order to smooth the data before and after revision we have imputed the rents of the year 2000 to the years before revision, taking inflation into account.

- Computations of the rental value of real estate are revised.
- Income on the individual level has limitations, because certain components (such as child benefits and rent subsidies) are ascribed to a different household member after revision (for example breadwinner instead of the head of the household). As this study focusses on total household income, this revision does not give us any problems.
- The method to determine whether persons on a certain address constitute a household together is changed
- Employer contribution payments are included in the wages before revision but are separated afterwards. To make the data comparable, we subtract employer contributions before revision.
- Contributions to social insurance, paid by the authority who pays out transfer incomes, were included before revision but were separated after revision. Unfortunately, it is not possible to subtract these contributions before revision, therefore we add them after revision to make the data comparable.
- After revision dividends from stocks of a substantial holding[20] also includes the selling of stocks from own business. Before revision these were excluded. Following the advice of Statistics Netherlands we try to exclude these dividends by dropping dividends which exceed 250,000 euro.
- Although we have tried to make definitions as consistent as possible, differences are left. Therefore, we smooth several income components on the individual level: labor income, occupational pensions, public pension benefits, and remaining transfer income (e.g. disability benefits and unemployment benefits). We have used absolute differences and take inflation into account.
- Sometimes the birth year of individuals in IPO are not consistent over time. For these people we have imputed the birth year from the population register (GBA), which is available as of 1995. There are a few observations that have inconsistent birth years and are not present in GBA (for example because they died before 1995).

[20] A taxpayer is regarded as having a substantial holding in a corporation if he or she, either alone or with his or her spouse, holds directly or indirectly 5% of the issued capital.

5.B The construction of income

This appendix describes our definition of income. Income is the sum of net non capital and net capital income. To construct net non capital income and net capital income we distinguish between the data before 2001 and after 2001, as in 2001 a new tax system was introduced. First we define net non capital and net capital income between 1989 and 2000. Secondly, we define net non capital and net capital income as from 2001. Net non capital income between 1989 and 2000 is defined by

$$\text{net non capital income} = L + T - \frac{L+T}{L+T+H+C}\tau_i - P + \text{allowances},$$

where L is the sum of all income obtained with labor, T is total transfer income, C is total capital income, τ_i is the total taxation on income (from labor, transfers, interests, etc.), and P is the sum of the forced premia for social security insurances and employees' insurances. In the Dutch law mortgage interests are tax-deductible. Furthermore, the law states that home owners earn a taxable income from an owner-occupied house (the so called 'imputed rent'). The imputed rent is a percentage of the value of the house determined by the municipal authority. In our calculations of the net capital and non capital income we take this deductible mortgage interests and imputed rent into account. H is the imputed rent minus the mortgage interests. Allowances consist of child benefits, rent subsidies etc. For the period 1989–2000 net capital income is defined by

$$\text{net capital income} = \left(H - \frac{H}{L+T+H+C}\tau_i\right) + \left(C - \frac{C}{L+T+H+C}\tau_i - \tau_w\right),$$

where τ_w is the tax on wealth. Net capital income consists of two parts. The first part is capital income associated with the possession of an own house, the second part is all remaining capital income. As from 2001 the tax system has changed, as from then we define non capital income as

$$\text{net non capital income} = L + T - \frac{L+T}{L+T+H}\tau_i - P + \text{allowances},$$

Compared to the period 1989–2000 C is no longer in the definition, because the taxation on income does not include the income on capital (interests, dividend etc.) anymore. As from 2001 we define capital income as

$$\text{net capital income} = \left(H - \frac{H}{L+T+H}\tau_i\right) + (C - \tau_w),$$

also here the definition has changed because the taxation on income does not include capital income anymore.

Estimation method 5.C

To estimate the parameter vector β of equation (5.1) we compute

$$y_{it} - \overline{y}_i = \beta\left(x_{it} - \overline{x}_i\right) + \left(v_{it} - \overline{v}_i\right), \tag{5.C.1}$$

where $\overline{y}_i$ is the average household income of household i across time. $\overline{x}_i$ and $\overline{v}_i$ are average values across time for each household i. Using (5.C.1), we can consistently estimate β with OLS. Furthermore, we correct the standard errors of β for the fact that the error terms $v_{it} - \overline{v}_i$ are correlated for observations of the same household (see for example Cameron and Trivedi, 2005, p. 727). For the computation of the variance-covariance matrix standard regression routines use $\sum_{i=1}^{N}(T_i) - K$ in the denominator of the multiplier, where T_i denotes the number of periods household i is in the data and K the number of explanatory variables. However, $\sum_{i=1}^{N}(T_i - 1) - K$ should be used here. Therefore, we multiply the variance-covariance matrix of the standard regression routine with $\left(\sum_{i=1}^{N}(T_i) - K\right) / \left(\sum_{i=1}^{N}(T_i - 1) - K\right)$ (see for example Baltagi, 1995, p.12).

Averaging (5.1) across all observations gives

$$\overline{y} = \alpha + \beta\overline{x} + \overline{v}, \tag{5.C.2}$$

when we use the identifying assumption $\sum_{i=1}^{N}\mu_i = 0$. From (5.C.2), we obtain $\widehat{\alpha}$ by

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}, \tag{5.C.3}$$

We also average the data of individual households across time

$$\overline{y}_i = \alpha + \beta \overline{x}_i + \mu_i + \overline{v}_i, \tag{5.C.4}$$

from (5.C.4) we compute $\widehat{\mu}_i = \overline{y}_i - \widehat{\alpha} - \widehat{\beta}\overline{x}_i$. Now $\widehat{v}_{it}$ is computed by

$$\widehat{v}_{it} = y_{it} - \widehat{\alpha} - \widehat{\beta} x_{it} - \widehat{\mu}_i. \tag{5.C.5}$$

which we use in the auxiliary regression for autocorrelation (5.3).

5.D Estimation results of the transition models

This appendix describes the estimation results of the transition models with regard to marital status, children, and labor market positions explained in section 5.4.3 and 5.4.4.

5.D.1 Marital status

Using GBA data we estimate transition models with regard to marital status. The explanatory variables are age, age squared, cohort, and cohort squared. We assume period effects to be negligible compared to the age and cohort effects. The first part of table 5.D.1 presents the results for the transition from being married to divorced. The probability of a divorce decreases with age for both men and women and as expected younger cohorts divorce more often than older cohorts. The second part of the table is about remarriage after a divorce. The probability to remarry after a divorce decreases with age and is smaller for younger cohorts than for older cohorts. Only for women, the probability to remarry increases up to the cohort born in 1946. The results for the transition from being unmarried to married show that the probability of marriage mostly decreases with age and that younger cohorts have a higher probability to marry than older cohorts. This may look strange, as it is commonly known that younger cohorts marry less often than older cohorts. However, this sign can be explained by younger cohorts marrying later in life than the older cohorts. Therefore, the probability of marriage in the age group under consideration

(36–90) is relatively high for young cohorts. The fourth part of table 5.D.1 is about remarriage after the death of a spouse. For widow(er)s the probability to remarry decreases with age and is higher for younger than for older cohorts.

Table 5.D.1: Logit transition models for marital status

	Men		Women	
	Coef.	S.e.	Coef.	S.e.
Married → Divorced				
age/10	1.461	0.0414	1.611	0.0482
$(\text{age}/10)^2$	-0.190	0.0043	-0.224	0.0052
(year of birth-1900)/10	0.179	0.0404	-0.097	0.0477
$((\text{year of birth-1900})/10)^2$	0.013	0.0037	0.038	0.0042
constant	-8.556	0.0894	-7.888	0.1054
pseudo R^2	0.050		0.056	
Divorced → Married				
age/10	-0.388	0.0560	0.167	0.0704
$(\text{age}/10)^2$	-0.032	0.0057	-0.098	0.0074
(year of birth-1900)/10	-0.068	0.0539	0.290	0.0708
$((\text{year of birth-1900})/10)^2$	-0.010	0.0050	-0.031	0.0064
constant	-0.069	0.1197	-2.975	0.1533
pseudo R^2	0.028		0.050	
Unmarried → Married				
age/10	-2.732	0.0751	-1.954	0.1066
$(\text{age}/10)^2$	0.203	0.0086	0.124	0.0123
(year of birth-1900)/10	1.654	0.0919	1.927	0.1305
$((\text{year of birth-1900})/10)^2$	-0.138	0.0077	-0.151	0.0108
constant	-0.918	0.1987	-4.154	0.2755
pseudo R^2	0.046		0.065	
Widowed → Married				
age/10	1.254	0.1353	1.234	0.1467
$(\text{age}/10)^2$	-0.175	0.0114	-0.188	0.0130
(year of birth-1900)/10	0.204	0.0919	0.419	0.1101
$((\text{year of birth-1900})/10)^2$	-0.026	0.0106	-0.018	0.0118
constant	-5.406	0.3319	-7.481	0.3376
pseudo R^2	0.092		0.143	

Children 5.D.2

Using IPO data we estimate the probability of children leaving their parental home from one year to the other. For this estimation we select all children in IPO in 2006 and check whether they are still in the household in 2007.

Thus, for the years 2008–2020 we assume children to have the same behavior with regard to leaving their parental home as the children between 2006 and 2007.[21]

Table 5.D.2: Logit transition models for children

	Coef.	S.e.
Children not leaving their parental home		
age	-0.008	0.0147
age^2	-0.005	0.0004
gender	0.474	0.0445
constant	4.072	0.1198
pseudo R^2	0.165	
N	38906	
New children being born		
age/10	-5.983	0.1449
(age/10)^2	0.459	0.0154
(year of birth-1900)/10	-0.432	0.1666
((year of birth-1900)/10)^2	0.073	0.0144
man	0.679	0.0232
couple	1.514	0.0473
one child	0.753	0.0287
two children	-0.637	0.0306
constant	11.220	0.4473
pseudo R^2	0.219	
N	850151	

As expected, the probability to leave the parental home increases with age. Furthermore, female children have a higher probability to leave their parental home than males.

Table 5.D.2 also presents the estimation results with regard to new children being born in a household. For this estimation we select all households in the years 1989–2006 and we determine, given the characteristics in $t-1$, whether a new child has entered the household during the next year. As from age 36 the probability of a new child decreases with age, and is higher for younger cohorts (who in general get children later in life). The probability of a new child is higher in households with a couple and in households where already one child is present. So, if there is already one child present, there is a relatively large probability of a second child after age 36. On the other hand, when there are

[21] Here, children are defined as all persons younger than 30 who are at least 18 years younger than the key person of a household.

already two or more children in the household, the probability of an extra child after age 36 is relatively low.

Labor market status 5.D.3

Transitions in labor market status are estimated for singles and couples separately. Table 5.D.3 shows the estimation results for singles. The first half of the table is about the transitions from work to occupational pension or to 'other' (being no labor and no occupational pension, for example the receivers of just unemployment benefits, disability benefits, or a state pension). The second half of the table deals with the transitions from 'other' to work or to occupational pension. The probability to keep on working in the labor market increases for younger generations. Furthermore, divorced men and women experience transitions from work to 'other' and from 'other' to work relatively often. For women, the number of children is positively associated with transitions from work to 'other' (e.g. out of the labor force).

The labor market outcomes of a couple are interrelated. We therefore treat the three possible outcomes of the two persons of a couple as nine (3×3) possible outcomes, and model the transitions between these nine states. The nine outcomes are listed in table 5.D.4, together with their relative frequencies. The table shows that the proportion of two-earner couples increased between 1989 and 2007, from 23% to 40%. On the other hand, the proportion of couples where only the man receives labor income declined (column three). Over the years, the percentage of couples involved in occupational pensions has increased. Especially the percentage of couples where both men and women receive labor income or occupational pension income has increased.

The explanatory variables used to explain transitions in the labor market positions of couples are the age of the man and woman part of the couple and the cohort to which the man belongs (in combination with the age of both man and woman, the cohort of the woman is automatically known). We add an interaction of age and cohort to allow for age patterns to be different for different cohorts.[22] Finally, table 5.D.5 shows the gender-specific

[22]To save space, the detailed estimation results with regard to the transitions between these nine states are available on request.

Table 5.D.3: Logit transition models for the labor market status of singles[a]

Men	Labor → Occup. pension		Labor → Other	
	Coef.	S.e.	Coef.	S.e.
age/10	0.234	0.6802	-1.194	0.2637
$(age/10)^2$	0.091	0.0457	0.114	0.0218
(year of birth-1900)/10	-0.139	0.2471	-0.314	0.0634
interaction age and cohort[b]	-0.020	0.0457	-0.009	0.0206
divorced	0.240	0.0802	0.303	0.0538
widow(er)	0.345	0.1096	-0.077	0.1227
constant	-6.315	2.3818	1.601	0.6178
pseudo R^2	0.063			
N	127759			
Women	**Labor → Occup. pension**		**Labor → Other**	
	Coef.	S.e.	Coef.	S.e.
age/10	0.121	0.7283	-1.001	0.2838
$(age/10)^2$	0.020	0.0476	0.090	0.0227
(year of birth-1900)/10	-0.812	0.2743	-0.341	0.0733
interaction age and cohort[b]	0.116	0.0488	-0.041	0.0218
divorced	-0.414	0.0800	0.454	0.0592
widow(er)	0.797	0.0805	0.048	0.1065
# children	-0.099	0.0659	0.137	0.0202
constant	-3.256	2.6181	1.954	0.7104
pseudo R^2	0.104			
N	83337			
Men	**Other → Labor**		**Other → Occup. pension**	
	Coef.	S.e.	Coef.	S.e.
age/10	-0.760	0.2741	2.067	0.8009
$(age/10)^2$	0.017	0.0227	-0.133	0.0513
(year of birth-1900)/10	0.139	0.0644	0.205	0.3114
interaction age and cohort[b]	-0.045	0.0219	-0.046	0.0524
divorced	0.502	0.0590	0.318	0.0907
widow(er)	0.219	0.1410	0.642	0.1074
constant	0.803	0.6192	-10.647	2.9932
pseudo R^2	0.183			
N	40821			
Women	**Other → Labor**		**Other → Occup. pension**	
	Coef.	S.e.	Coef.	S.e.
age/10	-0.486	0.2744	0.680	0.6978
$(age/10)^2$	-0.017	0.0222	-0.090	0.0426
(year of birth-1900)/10	0.254	0.0684	-0.882	0.2963
interaction age and cohort[b]	-0.065	0.0218	0.139	0.0456
divorced	0.582	0.0579	-0.222	0.0795
widow(er)	0.441	0.1036	0.960	0.0718
# children	-0.014	0.0169	-0.113	0.0546
constant	-0.164	0.6538	-3.676	2.7678
pseudo R^2	0.245			
N	49339			

[a] Estimation results of the transition models for labor market status. We assume 'occupational pension' to be an absorbing state.
[b] The interaction between age and cohort: age/10*(year of birth-1900)/10.

Table 5.D.4: Relative frequencies of the labor market positions of couples[a]

Year	1	2	3	4	5	6	7	8	9
1989	22.8	1.1	36.0	1.6	1.5	14.7	4.1	0.8	17.4
1990	24.2	1.2	34.1	1.9	1.6	15.7	4.1	0.8	16.5
1991	25.6	1.2	32.2	1.8	1.9	16.2	4.2	0.8	16.1
1992	26.3	1.2	30.5	1.8	2.0	17.2	4.5	0.8	15.7
1993	27.7	1.2	29.4	2.0	2.2	17.3	4.6	0.8	15.0
1994	28.1	1.2	28.3	2.1	2.4	17.6	4.8	0.8	14.7
1995	29.3	1.3	27.6	2.1	2.6	17.5	4.9	0.8	13.9
1996	30.4	1.4	26.7	2.1	2.9	17.6	4.9	0.8	13.2
1997	31.6	1.4	25.9	2.3	3.2	17.8	4.9	0.7	12.3
1998	33.4	1.5	24.8	3.4	2.9	16.9	4.6	0.7	11.8
1999	35.6	1.4	23.8	3.5	2.7	17.0	4.5	0.7	10.8
2000	36.9	1.7	22.0	3.4	3.4	17.3	4.5	0.7	10.1
2001	36.0	1.5	21.7	2.6	3.9	19.4	4.3	0.6	9.9
2002	38.4	1.8	19.9	2.7	4.3	18.0	4.9	0.7	9.2
2003	38.5	1.9	18.9	3.0	4.7	18.0	5.3	0.8	9.0
2004	38.8	2.0	18.2	3.3	5.1	18.0	5.2	0.8	8.6
2005	38.8	2.0	17.5	3.5	5.4	18.1	5.5	0.8	8.4
2006	39.8	2.0	16.4	3.8	6.0	18.1	5.5	0.8	7.7
2007	40.3	1.9	15.5	4.1	6.5	18.2	5.4	0.8	7.2

[a] We distinguish three labor market positions: (1) labor, (2) occupational pension, (3) other (none of these two). For couples we thus have 9 (= 3 × 3) combinations, corresponding to the nine columns in the table. 1: man works, woman works, 2: man works, woman occ. pension, 3: man works, woman other, 4: man occ. pension, woman works, 5: man occ. pension, woman occ. pension, 6: man occ. pension, woman other, 7: man other, woman works, 8: man other, woman occ. pension 9: man other, woman other.

estimation results of the multinomial logit model to determine the initial labor market status of new household members and children who enter adulthood. All household members in the data are used. For men and women the probability of labor increases until about age 40 and decreases afterwards. The probability of receiving an occupational pension, on the other hand, increases with age. Younger cohorts have a higher probability for a labor or occupational pensions status than older generations.

Table 5.D.5: Multinomial logit model for the initial labor market status of new household members and children entering adulthood

	Men		Women	
	Coef.	S.e.	Coef.	S.e.
Labor[a]				
age/10	8.778	0.0404	6.150	0.0371
$(age/10)^2$	-1.768	0.0105	-1.134	0.0095
$(age/10)^3$	0.107	0.0008	0.063	0.0008
(year of birth-1900)/10	0.207	0.0700	-1.925	0.0657
$(\text{year of birth-1900/10})^2$	0.074	0.0127	0.560	0.0115
$(\text{year of birth-1900/10})^3$	-0.007	0.0007	-0.036	0.0006
constant	-13.502	0.1349	-10.038	0.1377
Occupational pension				
age/10	-4.883	0.1267	0.545	0.1512
$(age/10)^2$	1.548	0.0221	0.248	0.0251
$(age/10)^3$	-0.104	0.0013	-0.020	0.0014
(year of birth-1900)/10	0.834	0.0690	0.064	0.0623
$(\text{year of birth-1900/10})^2$	-0.062	0.0157	0.076	0.0158
$(\text{year of birth-1900/10})^3$	-0.001	0.0012	-0.005	0.0013
constant	-5.856	0.2926	-10.375	0.3422
pseudo R^2	0.489		0.296	
N	1019127		998296	

[a] We distinguish three labor market positions: labor, occupational pension, and ‘other’ (no labor and no occupational pension income). ‘Other’ is the reference category in this estimation.

Extended estimation results 5.E

Table 5.E.1: Extended estimation results for Table 5.5.1, the age dummy coefficients[a]

	Coef 1	S.e.	Coef. 2	S.e.	Coef. 3	S.e.
age 36	-0.259	0.0050	-0.201	0.0051	-0.217	0.0055
age 37	-0.249	0.0049	-0.187	0.0050	-0.204	0.0054
age 38	-0.240	0.0048	-0.177	0.0050	-0.195	0.0054
age 39	-0.227	0.0047	-0.165	0.0049	-0.184	0.0053
age 40	-0.211	0.0047	-0.151	0.0049	-0.173	0.0053
age 41	-0.191	0.0046	-0.138	0.0048	-0.161	0.0052
age 42	-0.170	0.0045	-0.128	0.0047	-0.153	0.0051
age 43	-0.150	0.0045	-0.122	0.0046	-0.147	0.0051
age 44	-0.123	0.0044	-0.109	0.0046	-0.136	0.0050
age 45	-0.099	0.0043	-0.101	0.0045	-0.129	0.0049
age 46	-0.072	0.0043	-0.089	0.0044	-0.118	0.0048
age 47	-0.046	0.0042	-0.078	0.0043	-0.107	0.0048
age 48	-0.019	0.0042	-0.062	0.0042	-0.092	0.0047
age 49	0.006	0.0041	-0.045	0.0041	-0.075	0.0046
age 50	0.020	0.0041	-0.034	0.0040	-0.064	0.0045
age 51	0.031	0.0040	-0.025	0.0039	-0.053	0.0044
age 52	0.043	0.0039	-0.011	0.0039	-0.038	0.0043
age 53	0.046	0.0039	-0.005	0.0038	-0.031	0.0042
age 54	0.051	0.0038	0.005	0.0037	-0.018	0.0041
age 55	0.051	0.0037	0.010	0.0036	-0.011	0.0040
age 56	0.047	0.0036	0.011	0.0035	-0.005	0.0039
age 57	0.038	0.0036	0.008	0.0035	-0.005	0.0038
age 58	0.033	0.0035	0.007	0.0034	-0.002	0.0037
age 59	0.024	0.0034	0.003	0.0033	-0.002	0.0035
age 60	0.017	0.0033	-0.001	0.0032	-0.001	0.0033
age 61	0.008	0.0032	-0.006	0.0031	-0.001	0.0031
age 62	-0.003	0.0030	-0.013	0.0029	-0.005	0.0029
age 63	-0.007	0.0028	-0.013	0.0028	-0.004	0.0028
age 64	-0.008	0.0026	-0.012	0.0025	-0.002	0.0026
age 65[b]	0.000	-	0.000	-	0.000	-
age 66	-0.015	0.0024	-0.013	0.0023	-0.011	0.0023
age 67	-0.017	0.0026	-0.012	0.0025	-0.010	0.0025
age 68	-0.019	0.0028	-0.012	0.0027	-0.010	0.0026
age 69	-0.021	0.0029	-0.012	0.0028	-0.010	0.0028
age 70	-0.022	0.0030	-0.010	0.0029	-0.009	0.0029

[a] Estimation results continue on the next page.
[b] The reference category is ‘age 65’.

Table 5.E.1 continued

	Coef. 1	S.e.	Coef. 2	S.e.	Coef. 3	S.e.
age 71	-0.026	0.0032	-0.012	0.0031	-0.011	0.0030
age 72	-0.025	0.0033	-0.009	0.0032	-0.009	0.0031
age 73	-0.023	0.0033	-0.005	0.0032	-0.004	0.0032
age 74	-0.019	0.0034	0.000	0.0033	0.000	0.0033
age 75	-0.019	0.0036	0.002	0.0035	0.001	0.0034
age 76	-0.018	0.0037	0.005	0.0036	0.003	0.0036
age 77	-0.016	0.0039	0.008	0.0039	0.007	0.0038
age 78	-0.010	0.0041	0.016	0.0040	0.014	0.0040
age 79	-0.007	0.0043	0.020	0.0043	0.018	0.0042
age 80	-0.001	0.0045	0.028	0.0044	0.024	0.0043
age 81	0.002	0.0047	0.032	0.0047	0.028	0.0046
age 82	0.006	0.0051	0.038	0.0051	0.033	0.0050
age 83	0.004	0.0057	0.039	0.0056	0.033	0.0055
age 84	0.011	0.0062	0.048	0.0061	0.041	0.0060
age 85	0.018	0.0071	0.057	0.0069	0.048	0.0068
age 86	0.035	0.0079	0.075	0.0077	0.066	0.0076
age 87	0.031	0.0098	0.074	0.0096	0.063	0.0095
age 88	0.032	0.0131	0.074	0.0124	0.061	0.0122
age 89	0.050	0.0155	0.092	0.0146	0.081	0.0144
age 90	0.060	0.0222	0.099	0.0207	0.091	0.0211

Simulation results model specification 1 and 2 5.F

Table 5.F.1: Predictions of equivalized household income, age 50-64[a]

Year	Mean	p10	p50	p90	$\frac{p90}{p10}$	$\frac{p90}{p50}$	$\frac{p50}{p10}$	Gini
Model specification 1								
2008	24340	13167	21905	37813	2.87	1.73	1.66	0.244
2009	24418	13205	21996	37792	2.86	1.72	1.67	0.240
2010	24456	13261	22066	37849	2.85	1.72	1.66	0.240
2011	24471	13318	22110	37783	2.84	1.71	1.66	0.237
2012	24609	13437	22288	37814	2.81	1.70	1.66	0.238
2013	24815	13544	22400	38242	2.82	1.71	1.65	0.239
2014	24877	13511	22547	38347	2.84	1.70	1.67	0.240
2015	25146	13562	22600	38960	2.87	1.72	1.67	0.244
2016	25215	13619	22666	39207	2.88	1.73	1.66	0.242
2017	25362	13692	22787	39536	2.89	1.74	1.66	0.243
2018	25627	13760	22999	40067	2.91	1.74	1.67	0.245
2019	25940	13869	23163	40706	2.94	1.76	1.67	0.248
2020	26047	13893	23180	40909	2.94	1.76	1.67	0.251
Model specification 2								
2008	24659	13199	22205	38418	2.91	1.73	1.68	0.245
2009	24752	13276	22332	38160	2.87	1.71	1.68	0.242
2010	24950	13394	22558	38531	2.88	1.71	1.68	0.241
2011	25083	13601	22643	38561	2.84	1.70	1.66	0.240
2012	25153	13721	22857	38765	2.83	1.70	1.67	0.237
2013	25426	13688	22902	39381	2.88	1.72	1.67	0.241
2014	25564	13711	23081	39328	2.87	1.70	1.68	0.241
2015	25793	13828	23283	40008	2.89	1.72	1.68	0.242
2016	25890	13891	23238	40142	2.89	1.73	1.67	0.244
2017	26039	14051	23483	40459	2.88	1.72	1.67	0.242
2018	26125	13939	23528	40733	2.92	1.73	1.69	0.245
2019	26414	14026	23618	41364	2.95	1.75	1.68	0.248
2020	26516	14068	23688	41657	2.96	1.76	1.68	0.249

[a] In this study income is always inflated/deflated to 2005 euro's.

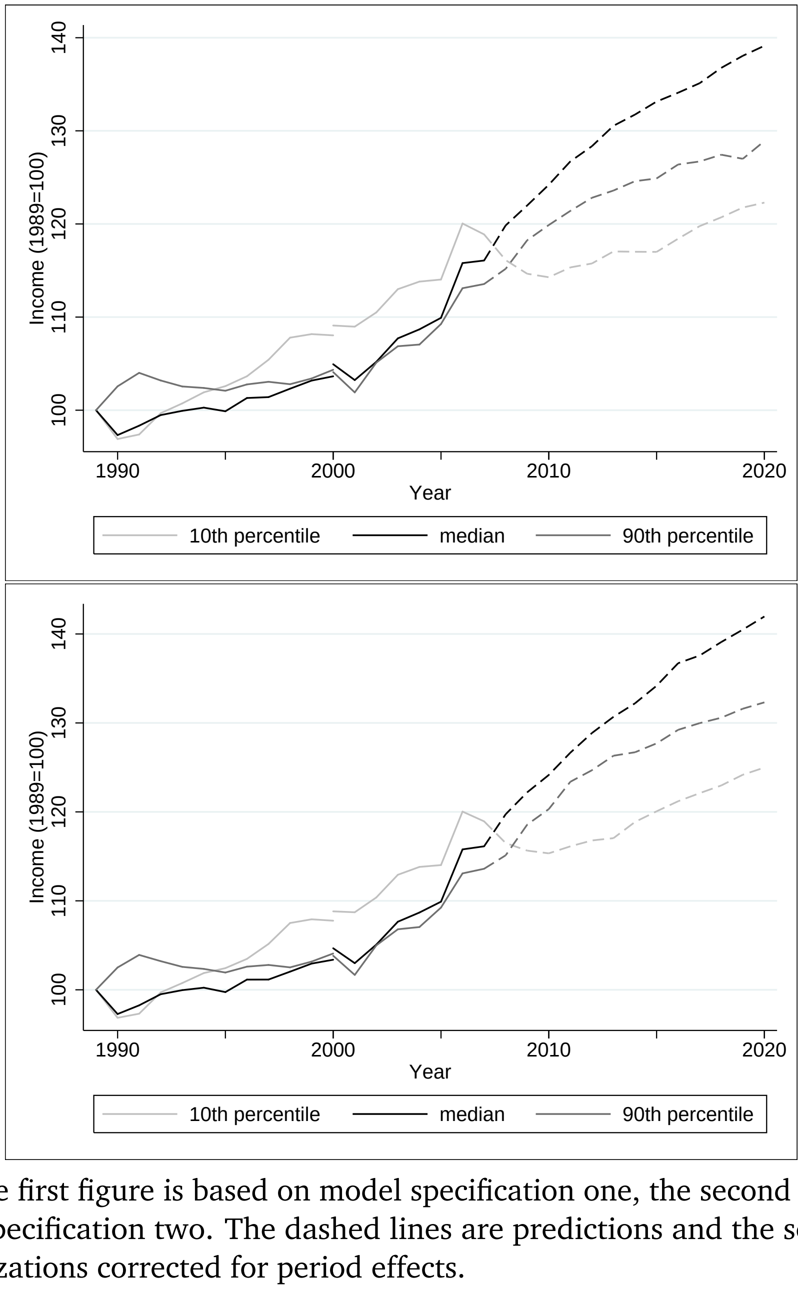
Table 5.F.2: Predictions of equivalized household income, age 65-90[a]

Year	Mean	p10	p50	p90	$\frac{p90}{p10}$	$\frac{p90}{p50}$	$\frac{p50}{p10}$	Gini
Model specification 1								
2008	20327	12178	17842	31186	2.56	1.75	1.47	0.227
2009	20634	12025	18162	32018	2.66	1.76	1.51	0.230
2010	20821	11985	18491	32459	2.71	1.76	1.54	0.231
2011	21159	12096	18866	32871	2.72	1.74	1.56	0.233
2012	21398	12140	19108	33252	2.74	1.74	1.57	0.232
2013	21646	12276	19435	33462	2.73	1.72	1.58	0.233
2014	21803	12272	19615	33734	2.75	1.72	1.60	0.231
2015	21975	12271	19822	33814	2.76	1.71	1.62	0.231
2016	22184	12419	19964	34220	2.76	1.71	1.61	0.230
2017	22365	12561	20116	34307	2.73	1.71	1.60	0.230
2018	22491	12660	20361	34504	2.73	1.69	1.61	0.228
2019	22635	12770	20554	34385	2.69	1.67	1.61	0.226
2020	22808	12826	20714	34901	2.72	1.68	1.62	0.228
Model specification 2								
2008	20315	12230	17843	31198	2.55	1.75	1.46	0.227
2009	20687	12140	18211	32134	2.65	1.76	1.50	0.231
2010	20925	12108	18501	32606	2.69	1.76	1.53	0.233
2011	21389	12191	18875	33440	2.74	1.77	1.55	0.236
2012	21707	12260	19203	33790	2.76	1.76	1.57	0.236
2013	21909	12287	19475	34233	2.79	1.76	1.59	0.236
2014	22188	12478	19702	34338	2.75	1.74	1.58	0.235
2015	22408	12604	19996	34609	2.75	1.73	1.59	0.234
2016	22644	12723	20373	35022	2.75	1.72	1.60	0.233
2017	22788	12819	20503	35225	2.75	1.72	1.60	0.231
2018	22954	12909	20728	35391	2.74	1.71	1.61	0.229
2019	23255	13035	20937	35665	2.74	1.70	1.61	0.231
2020	23425	13121	21156	35859	2.73	1.69	1.61	0.229

[a] In this study income is always inflated/deflated to 2005 euro's.

Indexed growth in specification 1 and 2 5.G

Figure 5.G.1: Indexed growth of equivalized household income in the age group 65-90.



Note: the first figure is based on model specification one, the second figure on model specification two. The dashed lines are predictions and the solid lines are realizations corrected for period effects.

5.H Theil decomposition

Overall income inequality can be related to mutually exclusive population subgroups using a Theil decomposition. We use the Theil decomposition method to explore whether the rising inequality in the lower part of the distribution is caused by an increase in the inequality between households with and without occupational pension income. Theil decompositions are developed by Shorrocks (1980), Bourguignon (1979), and Cowell (1980). The Theil index is a weighted average of inequality within subgroups, plus inequality among those subgroups. More specific, inequality within a year is the average inequality within each subgroup, weighted by the income of the subgroups, plus the inequality among subgroups. The subgroups in this study are (1) households with occupational pension income and (2) households without occupational pension income. The Theil index is given by

$$T = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\overline{y}}\log\left(\frac{y_i}{\overline{y}}\right) \tag{5.H.1}$$

where N is the number of observations, y_i the income of household i, and $\bar{y}$ average income of all households. (5.H.1) can be rewritten as

$$T = \left(s_1 T_1 + s_2 T_2\right) + \left(s_1 \log\left(\frac{\overline{y}_1}{\overline{y}}\right) + s_2 \log\left(\frac{\overline{y}_2}{\overline{y}}\right)\right) \tag{5.H.2}$$

where the first term presents within group inequality and the second term presents between group inequality. T_1 is the Theil index for households with occupational pension income, T_2 is the Theil index for households without occupational pension income. s_1 is the share of total income received by the households with occupational pension income, s_2 is the share of total income received by the households without occupational pension income. $\bar{y}_1$ is the average income of households with occupational pension income and $\bar{y}_2$ is the average income of households without occupational pension income.

6 The Association between Individual Income and Remaining Life Expectancy at the Age of 65 in the Netherlands

This chapter is based on Kalwij, Alessie, and Knoef (2009).

6.1 Introduction

Significant socioeconomic inequalities in mortality risk over many populations and time periods have been identified in the literature.[1] These socioeconomic inequalities in mortality, commonly termed differential mortality, give strong evidence for an inverse relationship between income and mortality risk. Estimates of this relationship for different populations with respect to country and age range indicate that the ratio of mortality risk for individuals in the lowest quartile of the income distribution to that of individuals in the highest quartile ranges from around two in Europe to three in the U.S.[2]

At the same time there is ongoing debate from various disciplines on the causal interpretation of this relationship and the possible pathways through which socioeconomic position affects health and mortality risk (e.g., Lindahl, 2005, Macintyre, 1997, Marmot et al., 1991, Smith, 1999, and Snyder and Evans, 2006). The explanation that differential mortality is a result of low-income individuals having access to less or lower quality health services appears

[1] See e.g., Attanasio and Hoynes (2000), Huisman et al. (2004), Hupfeld (2009), Hurd et al. (2001), Kunst et al. (2004), Marmot et al. (1991), Menchik (1993), Palme and Sandgren (2008), and Sullivan and Von Wachter (2009)

[2] See Duleep (1986) for the U.S., Martikainen et al. (2001) for Finland, Osler et al. (2002) for Denmark, Attanasio and Emmerson (2003) for the U.K., Blakely et al. (2004) for New Zealand, and Von Gaudecker and Scholz (2007) for Germany.

to be dismissed by the literature (see, e.g., Attanasio and Emmerson, 2003, and Smith, 1999). A cultural or behavioral explanation is that low-income individuals have an unhealthy lifestyle, e.g. are more often smokers, consume more alcohol and have diets that are linked to obesity (see, e.g., Macintyre, 1997, and Huisman et al., 2005). The influential Whitehall studies, however, have shown that behavioral risk is not the sole explanation for differential mortality (see, e.g., Marmot et al., 1991). This latter finding has sparked research that emphasizes long-term impacts on health of socioeconomic circumstances before adulthood and childhood health conditions (see, e.g., Barker, 1995, 1997, Case et al., 2005, 2002, and Van den Berg et al., 2006), cognitive abilities (Huisman and Mackenbach, 2007) and psychosocial reasons such as prolonged exposures to stress (e.g., Macintyre, 1997, Smith, 1999).

Independent of the causes of differential mortality, an inverse relationship between income and mortality risk has important implications for pension policy (e.g. Whitehouse and Zaidi, 2008).[3] Public pension policy in many countries, including the Netherlands, aims at redistributing income from the financially better to the financially worse off individuals. This redistribution may be adversely affected by differential mortality because, on average, low-income individuals receive public pension benefits for a relatively shorter period if the statutory retirement age is the same for all individuals (e.g., Nelissen, 1999). Likewise, for private pensions an inverse relationship between income and mortality risk implies that, compared to high-income individuals, low-income individuals' internal rate of return from a uniformly priced private pension plan is, on average, lower because their lower life expectancy results in receiving the benefits of the pension plan for a relatively shorter period (Bonenkamp, 2009, Brown, 2002, Hári, 2007, Menchik, 1993, Simonovits, 2006). These policy concerns are exemplified by the recent proposed pension reforms in the Netherlands that take explicitly into account this disparity by facilitating workers in low-income sectors to receive a pension at most two years before the (proposed) statutory retirement age of 67 (Stichting van de Arbeid, 2010).

To gain insights into the size of the above mentioned differences in life expectancy between low and high-income individuals, we empirically quantify

[3]For a discussion on the implications for public health policy we refer to epidemiological studies such as Huisman et al. (2005) and references therein.

the association between individual income and remaining life expectancy at 65, the statutory retirement age in the Netherlands.

The contribution of this study to the empirical literature on the relationship between income and mortality risk is threefold. First, in contrast to previous studies for the Netherlands that have most often examined the relation between education and mortality risk (e.g., Kunst and Mackenbach, 1994, Van Kippersluis et al., 2011), we estimate the association between income and mortality risk for individuals aged 65 or older. Moreover, we quantify the association between income and remaining life expectancy at age 65 by making use of Monte Carlo simulations. We choose age 65 because that is the statutory retirement age in the Netherlands after which all individuals receive (i) a public retirement pension that is independent of earnings history[4] and (ii), an occupational pension that depends on earnings history (see e.g., Nelissen, 1999). Because this retirement income consists primarily of pension income and is therefore closely related to individual earnings history, it serves as a good proxy for an individual's lifetime income.

Second, in contrast to most previous studies that, depending on data availability, have used either individual income or the sum of individual and spouse's income (i.e. household income), we make a distinction between individual income and spouse's income. This allows us to examine a frequent claim in the literature that material hardship, measured by household and not only individual financial resources, matter for health status and mortality risk (e.g., Martikainen et al., 2001). Indeed, for most women from the cohorts in our analysis, mortality risk is perhaps more likely to be negatively related to spouse's income than their own income because many have left the labor force at the time of marriage or birth of a first child. In the health stock (economic) model of Grossman (1972) an explanation for an association between spouse's socioeconomic position and individual mortality risk is that a spouse may improve the efficiency of an individual's investment in the health stock. The higher the partner's socioeconomic position, the greater the improvement and the lower mortality risk.[5]

[4] Eligibility and amount depend only on the years of recorded residency in the Netherlands between the ages of 15 and 65 (2% of the full public pension benefit for each year).

[5] In the model of Grossman (1972, 2000) health deteriorates with age at a relatively slower rate for individuals with a higher socioeconomic position which, in his model, implies that individual's socioeconomic position is positively related to both (lifetime) income and life expectancy.

A sociological explanation is that households have a shared lifestyle that is influenced by both partners (e.g., Torssander and Erikson, 2009). Our empirical model cannot discriminate between these different explanations but can test whether individual mortality for both men and women is associated with a spouse's income as well as the individual's income. Should we find that spouse's income is (negatively) associated with individual mortality risk, this could be interpreted as support for one of the above explanations. Limited empirical evidence is available on this issue and the exceptions being McDonough et al. (1999) for the U.S. and Torssander and Erikson (2009) for Sweden who report that for women, but not for men, spouse's income is negatively associated with individual mortality risk.

Third, our empirical model controls for unobserved individual-specific characteristics (i.e., random effects). This is of importance as it is inherent in the analysis of mortality risk that with age the sample becomes more selective in terms of both observed and unobserved characteristics. As a result, failing to control for this "dynamic selection" may bias the results (Cameron and Heckman, 1998, Van den Berg, 2001). Yet most of the studies cited above use (pooled) cross-sectional data, hence do not control for dynamic selection. Duration data make it possible to control for dynamic selection by including random effects. Nevertheless, using duration data, Hupfeld's (2009) model does not include random effects and Van den Berg et al. (2006) mention that including random effects does not affect the estimated impact of economic conditions early in life on individual mortality risk. Our empirical analysis uses panel data that are representative of the 65+ population and the model includes random effects to take into account dynamic selection. Nonetheless, a model complication does arise for individuals who enter the panel after age 65. That is, given the dynamic selection process, random effects at age 65 imply a dependency between these random effects and the covariates at later ages. To take this complication into account, we include dynamic sample selection correction terms that control for this dependency at the age of entry for individuals who enter the sample after age 65.

The chapter is organized as follows. Section 6.2 describes the data. Section 6.3 outlines the empirical model for analyzing mortality risk and explains the estimation procedure. Section 6.4 reports the analytical results, and section 6.5 summarizes the main findings and concludes the chapter.

Data 6.2

The data are taken from the 1996–2007 Income Panel Study of the Netherlands (IPO, Inkomens Panel Onderzoek, CBS 2009a) and the 1997–2008 Causes of Death registry (DO, Doodsoorzaken, CBS 2009b), both gathered by Statistics Netherlands. The IPO, a representative sample of the Dutch population, consists of an administrative panel dataset of about 92,000 individuals, randomly selected in 1996, which increased to about 99,000 individuals in 2007 because of population growth. Sampling is based on individuals' national security number, and the selected individuals are followed for as long as they are residing in the Netherlands on December 31 of the sample year. The dataset also includes individuals living in institutions for the elderly, such as nursing homes. Individuals born in the Netherlands enter the panel for the first time in the year of their birth; immigrants to the Netherlands, in the year of their arrival. An individual exits the panel on death or emigration from the Netherlands. Hence, the only reasons for panel attrition are mortality and emigration.[6]

The IPO contains data on the demographic characteristics and income of each member of a selected individual's household obtained from official institutions; most particularly, the population registry and tax office. The DO, on the other hand, provides information on date and cause of death for all residents deceased during the 1997–2008 period. These data come from medical records provided by medical examiners, who are legally obliged to submit them to Statistics Netherlands. The DO dataset also assigns a personal identifier that allows determination of whether an individual in the IPO has died by the next calendar year.

We select individuals aged 65 or over, who, because of population ageing, make up 12.8% of the sample in 1996 and 13.9% in 2007. This raw dataset consists of 151,120 observations for 21,159 individuals over the 1996–2007 period. We remove about 5% of the observations because of missing income information, about 1% because of missing values on marital status and 0.05% because of negative income. These exclusions affect relatively more men than

[6]The annual rate of attrition in survey data tends to be 20% or more (e.g., Attanasio and Emmerson, 2003), meaning that a strong relationship between attrition and, for instance, health status may lead to inconsistent estimates of the association between income and mortality risk. A further drawback of survey data is that individuals living in institutions for the elderly are often not included.

women (11% vs. 4%). Nonetheless, once we control for gender, the mortality rate among the individuals excluded is not significantly different from the mortality rate among those included, which suggests that these exclusions do not yield an endogenous sample selection.[7] Panel attrition for reasons other than mortality (i.e., because of emigration or missing values) is about 0.3% per year. The resulting sample consists of 19,258 individuals, 11,601 female and 7,657 male. Of these individuals, 42% enter the panel at the age of 65, while the remaining 58% enter the sample at a later age. The total sample contains 141,725 observations.

6.2.1 Variable definitions and descriptive statistics

The analysis is based on the variables gender, age, marital status, and income. We define age as the individual's age on January 1 of each year because in the Netherlands, the calendar year is also the fiscal year for income measurement, meaning that this choice ensures that income at age 65 is measured over the first entire calendar year of retirement. Table 6.2.1 and 6.2.2 report the number of observations by age and gender and the distribution of marital status by age and gender. The marital status variable distinguishes between a single adult household that includes divorcees (hereafter, 'single'), a married or cohabiting couple ('married'), and a widowed individual. The differences in marital status across age and gender result from the recognized fact that, on average, women live longer than men. These differences result in, for instance, an increasing proportion of women with age (last column, table 6.2.1) and, at a given age, relatively more widowed and fewer married women than men (table 6.2.2).

The IPO income data are based primarily on tax records and contain detailed and accurate information on all income components. Here, income is gross of income tax and social insurance contributions and is measured in 2005 euros using the consumer price index. Individual income is the sum of pension, labor, transfer, and capital income. Table 6.A.1 provides an overview of these components and their definitions for both men and women and shows that over 90% of retirement income is pension income. All income components are

[7] The p-value corresponding to the null hypothesis of no difference is equal to 0.261.

Table 6.2.1: Number of observations by age and gender

Age	Men	Women	Men (%)	Women (%)	Share of women (%)
65–69	17,504	23,349	32.0	26.8	57.2
70–74	15,943	21,558	29.2	24.7	57.5
75–79	11,530	18,593	21.1	21.3	61.7
80–84	6,305	13,307	11.5	15.3	67.9
85–89	2,577	7,085	4.7	8.1	73.3
90–94	664	2,692	1.2	3.1	80.2
95+	94	524	0.2	0.6	84.8
All	54,617	87,108	100.0	100.0	61.5

Table 6.2.2: Marital status by age and gender

	Marital status men			Marital status women		
Age	Single %	Widowed %	Married %	Single %	Widowed %	Married %
65–69	14.3	7.0	78.7	13.1	24.1	62.8
70–74	10.4	10.7	78.9	12.5	37.2	50.3
75–79	8.7	17.1	74.2	11.8	52.4	35.8
80–84	8.3	26.2	65.5	11.7	66.6	21.7
85–89	7.8	41.3	50.9	12.1	77.4	10.6
90–94	6.3	59.9	33.7	14.1	81.6	4.3
95+	4.3	62.8	33.0	16.0	83.0	1.0
All	10.9	14.8	74.4	12.4	46.4	41.2

observed for the individual and, in couple households, also for the spouse. The analysis excludes any income from other household members.[8]

As shown in table 6.2.3,[9] the mean income of single and widowed men is higher than that of single and widowed women, and the distribution of income for single and widowed men is wider than the distribution of income for single and widowed women. The income distribution for men shows a decrease in median income with age. This finding most likely results from changes in the income distribution over birth cohorts (Knoef et al., 2009). In addition, the

[8]About 8% of households have other household members, mostly children. Excluding these households would not affect the main results of this study. In addition, the main results are also unaffected by using a standardized (equivalized) income concept or using pension income only.

[9]Following Statistics Netherlands guidelines, we do not report statistics based on a number of observations below 25. These statistics are designated in the tables by "-".

Table 6.2.3: Distribution of individual income by marital status, age, and gender (euros)

	Single and widowed men				**Single and widowed women**			
Age	Mean	25th perc.	Median	75th perc.	Mean	25th perc.	Median	75th perc.
65–69	22,283	14,454	17,942	25,180	19,894	13,827	16,270	21,757
70–74	22,143	14,628	18,130	24,925	19,684	13,926	16,374	21,380
75–79	22,945	14,474	18,198	26,077	19,407	13,829	16,145	20,621
80–84	22,342	13,710	17,079	24,921	19,115	13,524	15,560	19,905
85–89	20,387	13,151	15,957	22,885	18,674	13,069	15,046	19,215
90–94	20,797	12,944	15,341	23,556	18,428	12,411	14,586	18,351
95+	18,210	12,868	15,213	20,000	19,055	12,203	14,169	18,249
All	22,163	14,194	17,640	25,045	19,345	13,597	15,843	20,518

	Married men				**Married women**			
Age	Mean	25th perc.	Median	75th perc.	Mean	25th perc.	Median	75th perc.
65–69	23,699	13,480	18,021	27,595	9,558	7,862	8,088	9,050
70–74	21,463	12,434	16,376	24,881	9,242	7,882	8,088	8,672
75–79	20,715	11,831	15,563	23,995	9,294	7,882	8,088	8,572
80–84	19,930	11,373	14,787	23,610	9,447	7,882	8,088	8,657
85–89	19,182	10,777	13,602	22,028	9,639	7,882	8,104	8,957
90–94	16,846	10,000	12,119	16,854	9,476	7,983	8,104	9,197
95+	22,604	9,984	12,702	32,184	-	-	-	-
All	21,809	12,391	16,469	25,324	9,407	7,882	8,089	8,818

income distribution for married women is rather compressed, partly because many retired married women have no earnings history[10] and receive only public pension benefits but also because pre-1990, part-time work often came with no pension plan or a pension plan that had a relatively high threshold before contributions could be made.

A comparison of the tables for married men and women reveals that women's income accounts for, on average, one-third of household income. The rank correlation between the individual's and the spouse's income is about 0.10 (not shown in a table).

[10] In our sample, about 25% of married women aged 65–69 receive an occupational pension; for women aged 70–74, this figure is about 20%, and for women aged 75 or over, it is about 16%.

Differential mortality 6.2.2

Defining mortality as being deceased in the following year, about 38% of the individuals died over the sample period. The results in table 6.2.4 confirm the accepted patterns that mortality risk increases with age and that men have a higher mortality risk than women. Likewise, age-specific mortality risk is lower among married individuals than among single or widowed individuals. In addition, the sample statistics on mortality risk by gender and age compare favorably with the population statistics from the Human Mortality Database (HMD column); that is, the differences between the two by age and gender are small.

Table 6.2.4: Mortality risk by gender, age, and marital status. The HMD columns provide population statistics from the Human Mortality Database[a] over the 1996–2006 period.

Mortality risk, men	Single	Widowed	Married	All	HMD
Age	%	%	%	%	%
65–69	4.6	3.4	1.5	2.1	2.2
70–74	5.8	5.5	3.5	3.9	3.8
75–79	8.3	8.6	6.1	6.7	6.4
80–84	13.6	13.0	9.8	11.0	10.8
85–89	23.8	18.5	15.9	17.6	17.5
90–94	33.3	30.2	24.1	28.3	27.5
95+	-	33.9	32.3	35.1	41.8
All	7.2	10.6	4.5	5.7	5.7

Mortality risk, women	Single	Widowed	Married	All	HMD
Age	%	%	%	%	%
65–69	1.6	1.7	1.0	1.3	1.2
70–74	2.7	2.5	1.6	2.1	2.0
75–79	4.9	4.3	3.0	3.9	3.5
80–84	8.2	8.1	6.8	7.9	6.6
85–89	12.8	13.6	10.8	13.2	12.4
90–94	20.3	20.3	19.8	20.3	21.7
95+	33.3	35.2	-	34.9	36.2
All	5.3	6.9	2.3	4.8	4.7

[a] Human Mortality Database, http://www.mortality.org.

Table 6.2.5: Mortality risk (%) by age, gender, and income quartile[a]

Panel A	**Single and widowed men** Income quartile					**Single and widowed women** Income quartile				
Age	Q1	Q2	Q3	Q4	Q1/Q4	Q1	Q2	Q3	Q4	Q1/Q4
65–69	5.5	4.2	3.4	3.8	1.4	2.7	1.2	1.4	1.4	1.9
70–74	6.6	7.5	5.2	3.5	1.9	3.8	2.5	2.4	1.8	2.1
75–79	12.0	8.2	6.9	7.3	1.6	6.2	4.7	3.5	3.5	1.8
80–84	16.4	12.1	12.2	11.4	1.4	10.1	7.7	7.1	7.5	1.3
85–89	22.1	19.1	18.8	15.6	1.4	17.8	11.3	11.2	12.1	1.5
90–94	35.7	30.8	27.0	23.5	1.5	22.3	22.4	17.9	16.3	1.4
95+	-	-	-	-	-	40.3	33.7	26.7	30.0	1.3
All	12.6	9.3	7.5	7.3	1.7	9.8	6.2	5.0	5.2	1.9
Panel B	**Married men** Income quartile					**Married women** Income quartile				
Age	Q1	Q2	Q3	Q4	Q1/Q4	Q1	Q2	Q3	Q4	Q1/Q4
65–69	3.2	1.8	1.0	1.0	3.3	1.5	1.0	0.8	0.7	2.1
70–74	4.6	3.6	3.3	2.3	2.0	2.3	1.2	1.5	1.3	1.7
75–79	7.2	6.5	5.4	4.6	1.6	3.4	2.9	2.5	3.2	1.1
80–84	11.3	10.0	8.5	8.2	1.4	7.6	7.2	6.9	5.3	1.4
85–89	17.1	13.4	16.5	15.5	1.1	11.8	10.1	9.5	11.8	1.0
90–94	28.3	20.4	21.4	14.7	1.9	34.8	14.8	6.3	26.5	1.3
95+	-	-	-	-	-	-	-	-	-	-
All	7.1	4.8	3.4	3.0	2.4	2.8	2.2	2.1	2.0	1.4
Panel C	**Married men** Spouse's income quartile					**Married women** Spouse's income quartile				
Age	Q1	Q2	Q3	Q4	Q1/Q4	Q1	Q2	Q3	Q4	Q1/Q4
65–69	1.8	1.3	1.6	1.4	1.3	1.3	1.1	1.1	0.8	1.7
70–74	4.3	3.1	3.1	3.3	1.3	2.1	1.1	1.7	1.6	1.3
75–79	7.2	6.4	5.1	5.5	1.3	2.9	3.1	3.6	2.7	1.1
80–84	11.8	9.2	8.9	9.6	1.2	8.2	7.5	5.5	5.5	1.5
85–89	16.2	17.4	15.5	14.5	1.1	13.2	6.1	11.0	11.0	1.2
90–94	25.4	32.5	15.6	26.3	1.0	27.5	16.7	19.4	10.7	2.6
95+	-	-	-	-	-	-	-	-	-	-
All	5.3	4.4	4.4	4.0	1.3	3.1	2.0	2.3	1.9	1.6

[a] The income quartiles are reported in table 2.

In table 6.2.5, panel A, we pool the results for single and widowed individuals as mortality risk by income quartile differs little between these two groups. Panel A shows that mortality risk decreases as income increases for single/widowed men and women, an effect that is strongest up to the third quartile. One measure of differential mortality is the ratio of mortality risk among individuals in the first quartile of the income distribution to mortality risk among individuals in the fourth quartile (columns Q1/Q4).

A comparison of panels A and B reveals that differential mortality is stronger for married men than for single/widowed men (2.4 vs. 1.7) but weaker for married women than for single/widowed women (1.4 vs. 1.9). These statistics are in line with findings of other European studies but lower than those reported for the U.S. (see the introduction). Overall, differential mortality appears to decline with age. Interestingly however, as illustrated in panel C, when spouse's income is considered, a differential mortality pattern emerges for both men and women, one that, although weaker for men, is similar to that produced for individual income. We are not aware of any comparable statistics in the literature.

Mortality risk model 6.3

This section outlines our empirical model for analyzing mortality risk. As discussed in section 6.2, we observe whether or not an individual is deceased in the subsequent year. Specifically, we consider the following latent variable model that relates next year's mortality risk, at age $(a+1)$, to individual's characteristics at age a

$$H_{a+1} = -\alpha_a - X_a\beta - \Lambda - \varepsilon_a, \quad \begin{cases} M_{a+1} = 1 \text{ if } H_{a+1} < 0 \\ M_{a+1} = 0 \text{ otherwise} \end{cases} \tag{6.3.1}$$

In the context of this study, and in line with Grossman's model, the latent variable H_{a+1} can be thought of as an individual's stock of health. If next year this stock falls below a certain threshold, normalized to zero in equation (6.3.1), the individual is deceased. The variable M_{a+1} denotes observed mortality at age $(a+1)$, and M_{a+1} is equal to one if an individual became a years old and died at age $(a+1)$, and zero otherwise. α_a is an age-specific intercept, while X_a is a $(1 \times k)$ vector of an individual's observed characteristics at age a, including marital status and income, with a corresponding $(k \times 1)$ parameter vector β. Λ denotes an individual's unobserved characteristics, i.e. a random effect, and is assumed to be constant over time and independent of the covariates at age 65. This assumption does not, however, exclude dependency between the covariates and Λ at later ages. We also assume that the random effect is normally distributed with a zero mean and σ^2_Λ variance and that the error term ε_a follows a logistic distribution and is independently distributed across individuals and time with a zero mean and a variance normalized to $\pi^2/3$.

6.3.1 Dynamic selection

Inherent in any study of mortality risk over the lifecycle is recognition that the population at risk changes with age. By explicitly accounting for random effects, our model allows for sample selection by age on the basis of both observed and unobserved characteristics. Not accounting for such dynamic selection may yield inconsistent estimates of α_a and β (Cameron and Heckman, 1998, Van den Berg, 2001). If mortality risk is negatively related to income and positively to unobserved characteristic(s) (Λ), low-income (high-income) individuals with a low (high) Λ value are more likely to survive another year than low-income individuals with a high Λ value. Hence, in this example, dynamic selection results in a population at risk in which the correlation between Λ and income becomes increasingly positive with age up to a certain age and then decreases thereafter as the sample becomes more homogenous with respect to income and Λ.

When all individuals are observed from the age of 65, the model takes dynamic selection into account. However, as reported in section 6.2, 58% of the individuals enter the sample after age 65 and dynamic selection implies that these individuals, having survived from age 65 to the age of panel entry (τ), are a selective sample (of their cohort) in terms of their covariates (including income) and Λ. A post-65 entry thus produces the empirical complication of a dependency between the random effect and the covariates, a dependency that, as already explained, changes with the age of entry. Ideally one would like to control for this by explicitly modelling the probability of survival up to the age of entry (see, e.g., Ridder, 1984) but this would require data on the covariates from the age of 65 up to the age one year before entry, which we do not have. The solution we propose here is to explicitly account for the change in the dependency between random effects and income at the age of entry, thereby maintaining the random effects assumption at age 65. Essentially, using the formula below, we parameterize the change in the dependency between the covariates and the random effect Λ at the age of first observation τ for individuals who enter the sample after the statutory retirement age of 65

$$\Lambda = \widetilde{X}_{\tau}\gamma + \theta \tag{6.3.2}$$

where $\widetilde{X}_{\tau} = \left((1, X_{\tau}) \times (\tau - 65), (1, X_{\tau}) \times (\tau - 65)^2\right)$ with a corresponding $((2k+2) \times 1)$ parameter vector γ. Scaling the effects of the covariates at age τ with the factors $(\tau - 65)$ and $(\tau - 65)^2$ takes into account, for instance, that the dependency between the random effect and income becomes more positive with age and at some age decreases as the sample becomes more homogeneous. θ is a random effect that is assumed to be independent of $\widetilde{X}_{\tau}$ and normally distributed with a zero mean and σ^2 variance. Consistent estimates of the α_a's and β are obtained under the additional assumption formalized in equation (6.3.2). Testing the joint hypothesis $\sigma^2 = 0 \cap \gamma = 0$ tests for the presence of random effects and the implied need to include dynamic sample selection correction terms (i.e., the additional covariates $\widetilde{X}_{\tau}$).

Estimation, identification, and empirical specification 6.3.2

Given the model outlined above, age-specific mortality risk conditional on observed and unobserved characteristics can be written as follows

$$Pr\left(M_{a+1} = 1 \mid X_{\tau}, X_a, M_a = 0\right) = F\left(\alpha_a + X_a\beta + \widetilde{X}_{\tau}\gamma + \theta\right) \tag{6.3.3}$$

The condition $M_a = 0$ formalizes the fact that all individuals in the population at risk are alive at age a, and $F(.)$ is the logistic cumulative distribution function. Equation (6.3.3) is the basis for a likelihood function in which we integrate out the unobserved individual specific effect. With i denoting the individual, $\tau(i)$ is the age of the individual when first observed in the sample and $A(i)$ is the age of the individual when last observed in the sample. The variable $m(i)$ is equal to one if the individual is deceased by age $A(i) + 1$, and zero otherwise. Maximum likelihood estimates of the model parameters are given by

$$
\begin{aligned}
(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\sigma}) = \underset{\alpha,\beta,\gamma,\sigma}{\operatorname{argmax}} \sum_{i=1}^{N} \\
\log \left(\int_{-\infty}^{+\infty} \left(\prod_{a=\tau_i}^{A(i)-1} \left(1 - F\left(\alpha_a + X_a(i)\beta + \widetilde{X}_{\tau(i)}(i)\gamma + \theta \right) \right) \right)^{I(A(i)>\tau(i))} \right. \\
\times \left(1 - F\left(\alpha_a + X_{A(i)}(i)\beta + \widetilde{X}_{\tau(i)}(i)\gamma + \theta \right) \right)^{1-m(i)} \\
\left. \times \left(F\left(\alpha_a + X_{A(i)}(i)\beta + \widetilde{X}_{\tau(i)}(i)\gamma + \theta \right) \right)^{m(i)} d\Phi\left(\frac{\theta}{\sigma} \right) \right), \qquad (6.3.4)
\end{aligned}
$$

where $\alpha = (\alpha_{65}, ..., \alpha_T)$,[11] N is the number of individuals, and Φ is the cumulative normal distribution function. The estimated model is often referred to as a random effects panel data logit model (Wooldridge, 2001). Equation (6.3.4) imposes proportionality between the age pattern, the covariates, and the random effect to ensure identification of the random effects distribution (see e.g., Cameron and Trivedi, 2005). For individuals who enter the sample at age 65 no dynamic sample selection correction terms are included (as for these $\tau = 65$, hence $\widetilde{X}_{\tau(i)}$ is a vector with zeros) and identification of the γ parameters is solely established by having individuals who enter the sample after age 65. We estimate the model separately for men and women.[12]

To estimate the association between the individual and spouse's income and mortality risk, we parameterize equation (6.3.1) as follows

$$
\begin{aligned}
-H_{a+1} =& \alpha_0 + \alpha_1 \mathrm{AGE}_a + \beta_1 \mathrm{MARRIED}_a + \beta_2 \mathrm{WIDOW}_a + \beta_3 Y_a^I \\
&+ \beta_4 Y_a^P + \beta_5 \left(Y_a^I\right)^2 + \beta_6 \left(Y_a^P\right)^2 + \beta_7 \left(Y_a^I Y_a^P\right) + \Lambda + \varepsilon_a. \quad (6.3.5)
\end{aligned}
$$

The associations between mortality risk and individual income (Y_a^I) and spouse's income (Y_a^P) are given by the parameters $\beta_3 - \beta_7$, and spouse's income is equal to zero for a single or widowed individual. The main advantage of this specification is that it nests two empirical specifications used in the literature and discussed in the introduction. The first nested model, which refers to the hypothesis $\beta_4 = \beta_6 = \beta_7 = 0$, is that in which only individual (and not spouse's) income has a direct association with mortality risk. The second nested model,

[11] T is the maximum age an individual may reach.
[12] We use the *xtlogit* command of the software package STATA (www.stata.com) for estimation.

which refers to the hypothesis $\beta_3 = \beta_4 \cap \beta_5 = \beta_6 \cap \beta_7 = 2\beta_5$, is that in which household income rather than individual and/or spouse's income is associated with mortality risk.

MARRIED is a dummy variable equal to one if the individual is married (including cohabitation) and zero otherwise; WIDOW is a dummy variable equal to one if the individual is widowed and zero otherwise. The reference category for marital status is a single adult household. Besides controlling for time effects, we also test for age-specific intercepts instead of a linear age function – which is briefly discussed in the next section.

Monte Carlo simulations 6.3.3

Whereas the parameter estimates of the model outlined above provide insights into the direction and relative size of the associations between the covariates and mortality risk, they offer no clear insights into the quantitative association with remaining life expectancy at age 65. Therefore, the second part of the empirical analysis simulates remaining life expectancy at 65 and examines how it is associated with individual and spouse's income by gender and marital status at the age of 65. The model can thus be seen as a period-age model that conforms to the life tables used by Statistics Netherlands to calculate life expectancy based on the age-specific mortality risks of the current population (Van der Meulen and Janssen, 2007).[13] We use the estimation results of the model outlined in sections 6.3.1 and 6.3.2 to simulate remaining life expectancy at age 65 (see appendix 6.B for the technical details of the Monte Carlo simulations).

Although income during retirement depends on lifetime earnings, the way that these two relate depends on the rules of the public and private pension systems. Hence, in the simulations, we take these rules into account when calculating income during retirement conditional on income before retirement (also known as pension-related gross yearly salary), reported in table 6.4.1 for several types of households.

As explained earlier, the occupational pension income an individual receives in addition to the public pension is dependent on this gross pension-related

[13]Theoretically, a cohort approach represents an alternative methodology. However, the available data do not allow for this method because, ideally, all individuals in the chosen cohorts should be followed up to the time of death.

salary. In the simulations we assume that individuals have lived in the Netherlands from the age of 15 onward and are therefore entitled to a flat rate public pension benefit that depends only on household composition. We also assume that, in the case of a married couple, the spouse's age is the same as the age of the individual. The baseline cases are individuals with a pension-related gross yearly salary equal to the median income in the Netherlands in 2005, which is 29,500 euros for full-time employees and 14,750 euros for part-time workers.[14] We further assume that the occupational pension is based on 40 years of employment. In the cohorts included in the analysis, most men worked full time before the age of 65, but most women either did not work (or are not entitled to occupational pension benefits) or worked mostly part time (about a quarter) before the age of 65. Not only do the household types in table 6.4.1 take these situations into account, but even when other household types or different assumptions of spouse's age, employment history, and part-time earnings are considered, the main simulation results for deviations from the baseline are rather insensitive to (minor) changes in the baseline cases.

The rules we apply for calculating income during retirement conform to the rules applied by the largest pension funds (see the footnote to table 6.4.1). For instance, they take into account that when a woman becomes widowed, she is entitled to a part of the deceased husband's occupational pension. In the simulation exercise, for each household type, we consider differences from the baseline situation that result from changes in the pension-related gross yearly salary. Hence, in the simulation results, median income refers to median pension-related gross salary (see table 6.4.1), low income refers to a pension-related gross salary based on minimum wage throughout the working life, and high income refers to a pension-related gross salary based on twice the median income (see table 6.A.2, appendix 6.A). These income classifications correspond roughly to the averages in the lowest and highest income quartiles for the different household types.

6.4 Empirical results

The estimation results for the model outlined in section 6.3 are reported in table 6.4.2 and discussed in section 6.4.1. The simulation results are given in

[14] Source: https://statline.cbs.nl (Statistics Netherlands). We assume that a part-timer works half time with a median income equal to 50% of that of a full-time worker.

Table 6.4.1: Income during retirement for different household types (euros)

Cells: Yearly amounts in 2005 euros	
Statutory minimum wage, full time	16392
Public pension for a single person household	11705
Public pension per person for a two person household	8018
	Baseline situation
Pension-related gross yearly salary, median[a]	
Man (full time)	29500
Woman (part time)	14750
Annual income during retirement by household type[b]	
Single person household	
Man (employed full time before age 65)	20650
Woman (employed part time before age 65)	16178
Couple, before age 65, the man was employed full time and the woman was not employed	
Man married at age 65	16962
Woman married at age 65	8018
Household income while married	24980
Man is widowed	20650
Woman is widowed	18094
Couple, before age 65, the man was employed full time and the woman part time	
Man married at age 65	16962
Woman married at age 65	12490
Household income while married	29452
Man is widowed	23845
Woman is widowed	22567

[a] Source: https://statline.cbs.nl, Statistics Netherlands. The (flat) rate public pension depends only on household composition and years of resident in the Netherlands. Pension-related gross salary refers to the salary on which the occupational pension is based and that is received in addition to a public pension.

[b] In these calculations we assume that (i) men worked full-time and women, if worked, worked parttime before age 65, (ii) a part-timer works half time with a median income equal to 50% of that of a full-time worker, (iii) individuals have lived in the Netherlands from the age of 15 onward, (iv) in the case of a married couple, the spouse's age is the same as the age of the individual, (v) the occupational pension in based on 40 years of employment. The rules we apply are roughly conform to those of the largest pension funds in the Netherlands and take into account widowhood. Occupational pension income is calculated based on pension-related gross salary (Y) and using the formula $\max(0, 0.7 \times \text{FTE} \times (Y - 10/7 \times \text{franchise}))$. A widowed individual receives $\max(0, 5/7 \times 0.7 \times \text{FTE} \times (Y_p - 10/7 \times \text{franchise}))$, where Y_p refers to the pension-related gross salary of the deceased spouse. The franchise is set equal to 11705 euro (the public retirement pension received by a person living alone). FTE is equal to one if the individual was employed full-time and 0.5 if employed part-time.

table 6.4.3 and discussed in section 6.4.2. In the following discussion we use a 5% level of significance. It is first worth noting, however, that we find no nonlinear age effects (in the index). We have examined this aspect using a model that contains a nonparametric specification of the age dependency of mortality risk and statistical tests have revealed that for both men and women, the nonparametric age function can be restricted to a linear function of age.[15]

6.4.1 Estimation results

An interpretation of the estimated associations in table 6.4.2 between marital status and mortality risk requires taking into account spouse's income, which is at least equal to the public pension benefit. We return to this in section 6.4.2.

As shown in table 6.4.2, individual income is negatively associated with mortality risk for both men and women and the parameter estimates are roughly of the same magnitude. Based on these estimates, we calculate that the negative association between income and mortality risk diminishes after about 130,000 euros (which is around the top 0.2 percentile of the income distribution). Moreover, although the negative association between mortality risk and spouse's income is insignificant for men and only significant at a 10% level for women (see the second last row of table 6.4.2), it is of roughly the same magnitude for both. In addition, the test statistics for men and women indicate that we do not reject the hypothesis that it is household income rather than individual and/or spouse's income that is associated with mortality risk (see the last row of table 6.4.2).

The estimate of the standard deviation of the random effect is significant for both men and women, which suggests the presence of unobserved individual-specific heterogeneity (random effects). The standard deviations are, however, relatively small compared to the standard deviation of the error term in equation (6.3.1), which is equal to $\sqrt{\pi^2/3} \approx 1.81$. As shown in table 6.4.2, we also test for the importance of controlling for random effects and the implied dynamic sample selection correction terms, as formulated in equation (6.3.2). These tests suggest a rejection of the null-hypothesis of no random effects (and

[15]See appendix 6.A. The p-values corresponding to an LR-test of model 2 against model 1 in table 6.A.3 and of model 7 against model 6 in table 6.A.4 are 0.33 and 0.18, respectively.

Table 6.4.2: Estimation results[a]

Dependent variable: Mortality risk		Men		Women	
Covariate, parameter, Eq. (6.3.5)		parameter estimate	S.e.	parameter estimate	S.e.
Constant	α_0	-11.636	1.054	-11.337	0.891
Age	α_1	0.132	0.016	0.109	0.014
Single		0.000		0.000	
Married	β_1	-0.872	0.142	-0.352	0.136
Widowed	β_2	-0.436	0.104	0.014	0.107
(Individual income/10,000)	β_3	-0.214	0.041	-0.271	0.048
(Spouse's income/10,000)	β_4	-0.097	0.124	-0.103	0.056
(Individual income/10,000)^2	β_5	0.008	0.002	0.010	0.003
(Spouse's income/10,000)^2	β_6	-0.003	0.015	0.002	0.004
(Individual income/10,000) × (Spouse's income/10,000)	β_7	0.028	0.016	0.015	0.022
Standard deviation random effect	σ	0.365	0.176	0.233	0.047
Log-likelihood value		-10888.9		-14823.3	
Number of observations		54617		87108	
Number of individuals		7657		11601	
Number of parameters		36		36	
Test of hypothesis		p-value		p-value	
Random effects and dynamic sample selection correction terms[b]		0.000		0.090	
Individual income[c]		0.000		0.000	
Spouse's income[d]		0.554		0.091	
Household income vs. individual/spouse's income[e]		0.709		0.356	

[a] The estimates of all model parameters are reported in tables A3 (model 2) and A4 (model 7), appendix 6.A.
[b] $H_0 : \sigma^2 = 0 \cap \gamma = 0$ (equation 6.3.2)
[c] $H_0 : \beta_3 = \beta_5 = 0$
[d] $H_0 : \beta_4 = \beta_6 = \beta_7 = 0$
[e] $H_0 : \beta_3 = \beta_4, \beta_5 = \beta_6, \beta_7 = 2\beta_5$

the implied correction terms) for men, but for women only at a 10% level of significance. Tables 6.A.3 (model 5) and 6.A.4 (model 10), appendix 6.A, report the estimation results for the model that excludes random effects and dynamic sample selection correction terms. We do not discuss these in detail but conclude, in line with Van den Berg et al. (2006), that overall the differences with the estimates of table 6 are relatively small.

6.4.2 Simulation results

The Monte Carlo simulations are based on the estimation results in table 6.4.2. These simulations, whose results are given in table 6.4.3, quantify the association between income and remaining life expectancy at the age of 65 in 2005 by gender for the different household types and income scenarios (see tables 6.4.1 and 6.A.2, appendix 6.A). The point estimates of (unconditional) remaining life expectancy at age 65 are 16.0 years for men and 18.4 for women.[16]

As discussed in section 6.3, the baseline situation is for median income individuals and considers three types of households. For single men, remaining life expectancy is almost 12 years and for single women about 18 years (first column, first two rows of table 6.4.3). Remaining life expectancy at 65 is considerably higher for married individuals, about 16 years for married men (first column, third row) and 20 years for married women who have been employed part time (first column, last row). The estimated differences between single and married individuals are significant: 4.4 years for men and 2.0 years for women.[17]

Next, we consider for each household type differences from the baseline situation that result from changes in income. The second and third columns report on the change in remaining life expectancy associated with a 10% higher than median income for either the man or the woman in each of the three household types. As table 6.4.3 shows, this association differs little across gender and household type: the point estimates vary between 0.16 and 0.20 with standard errors equal to 0.04. These estimates imply that life expectancy at age 65 for individuals with a 10% above median income is about two to two-and-a-half months higher than that for individuals with median income.

In addition, the fourth and fifth columns report the differences in remaining life expectancy between low and median income individuals for the different household types. An individual on minimum wage or with no earnings (low income) during the working life receives only a public retirement pension benefit during retirement. For both men and women, and irrespective of marital status at age 65, the difference in remaining life expectancy is less

[16] For this purpose, we estimate the mortality risk model using only age and year dummies. The simulation-based estimates for remaining life expectancy are 15.99 (0.49) for men and 18.41 (0.38) for women.

[17] Estimates of these differences are 4.396 (0.515) for men and 2.032 (0.774) for women.

than one year (fourth and fifth columns). The differences in remaining life expectancy between high-income individuals, those who earned twice the median income during the working life, and median income individuals are reported in the sixth and seventh columns. For men and women, the point estimates of these differences fall between 1.53 (for married women) and 1.91 (for single men). A comparison of the differences between these two extremes reveals that the difference in remaining life expectancy at age 65 between low-income individuals with only a public retirement pension and high-income individuals with twice the median income is about two-and-a-half years for both men and women.

Finally, we turn to the association between remaining life expectancy at 65 and spouse's income. As shown in the fourth and sixth rows of table 6.4.3, compared to married women whose spouse earned median income, remaining life expectancy at age 65 of women whose spouse earned a 10% above the median income is 0.18-0.19 years higher (about two-and-a-half months). For married men (fifth line), however, this difference is small and insignificant. In contrast, the difference in remaining life expectancy at 65 between women with a low-income spouse and women with a high- income spouse (see columns four and six, last row) is more than two years (1.62 - (-0.79)). This difference is explainable in two ways. First, spouse's income is negatively associated with women's mortality risk, albeit only at a 10% level of significance (see table 6.4.2). Second, and more important, marriage is negatively associated with mortality risk and women benefit, on average, from an extended duration of marriage when married to a high-income man. For men, however, we find no such indirect association because, having on average a shorter life span than women, men benefit relatively less from a spouse's higher income.

Table 6.4.3: Simulations of remaining life expectancy at the statutory retirement age of 65 (in years) by gender and household type[a]

	Baseline situation	Difference from the baseline situation due to differences in lifetime income					
	Median income	**Man Median+10%**	**Woman Median+10%**	**Man Low income**	**Woman Low income**	**Man High income**	**Woman High income**
Type of household	*Remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*
Single person household							
Man	11.73 (0.66)	0.20 (0.04)		-0.88 (0.17)		1.91 (0.39)	
Woman	18.18 (1.14)		0.18 (0.04)		-0.80 (0.17)		1.74 (0.36)
Couple, before age 65, the man was employed full time and the woman was not employed							
Man	16.12 (0.57)	0.20 (0.04)		-0.91 (0.18)		1.86 (0.38)	
Woman	19.50 (0.96)	0.19 (0.06)		-0.82 (0.23)		1.71 (0.50)	
Couple, before age 65, the man was employed full time and the woman part time							
Man	16.35 (0.62)	0.19 (0.04)	0.05 (0.05)	-0.85 (0.18)	-0.22 (0.21)	1.72 (0.39)	0.50 (0.66)
Woman	20.21 (1.01)	0.18 (0.05)	0.16 (0.04)	-0.79 (0.23)	-0.71 (0.17)	1.62 (0.47)	1.53 (0.35)

[a] Based on the estimation results of table 6.4.2. Tables 6.4.1 and 6.A.2 provide details on the income classifications. Standard errors are in parentheses.

Summary 6.5

This analysis quantifies the association between individuals' income and remaining life expectancy at the statutory retirement age in the Netherlands of 65. For this purpose, we estimate a mortality risk model that explicitly controls for unobserved individual-specific heterogeneity (random effects) using administrative data taken from the 1996–2007 Income Panel Study of the Netherlands supplemented with data from the Causes of Death registry.

Our main empirical findings are threefold. First, concerning model specification we find a significant presence of unobserved individual-specific heterogeneity (random effects) for men and for women only at a 10% level of significance. In our application the effects on the estimates when excluding random effects and the required dynamic sample selection correction terms are relatively small and, in other words, random effects play only a minor role in our mortality risk model.[18] Nevertheless, our findings underscore the importance of controlling for random effects and dynamic sample selection correction terms if inconsistent estimates are to be avoided.

Second, concerning the association between spouse's income and individual mortality risk we find that, conditional on marital status, spouse's income is only weakly (at a 10% level) associated with mortality risk for women. The literature suggests that an association with spouse's income might exist if what matters for health or mortality risk is material hardship, measured by household and not only individual financial resources (e.g., Martikainen et al., 2001) or if couples have a shared lifestyle that is influenced by both partners (e.g., Torssander and Erikson, 2009). Our findings do not provide strong support for such explanations. This perhaps surprising conclusion could be interpreted as evidence that early life health and socioeconomic circumstances (before marriage) are important contributors to later life differential mortality (see, e.g., Case et al., 2005, 2002). This suggests that public health policy such as (indirect) subsidized and universal health care for children and (means-tested) child allowances may reduce later life differential mortality.

Third, concerning the association of individual income with mortality risk we find that individual income is about equally strong and negatively asso-

[18]For completeness, table 6.A.5 (appendix 6.A) reports the simulation results based on a mortality model that does not control for random effects and dynamic sample selection. The differences with table 6.4.3 are also small.

ciated with mortality risk for men and women. The difference in remaining life expectancy at age 65 between low-income individuals with only a public retirement pension and high-income individuals with twice the median income, is about two-and-a-half years for both men and women. As discussed in the introduction, public pension policy in many countries, including the Netherlands, aims at redistributing income from the financially better to the financially worse off individuals, a redistribution that is adversely affected by this difference because low-income income individuals, on average, receive public pension benefits for a relatively shorter period. Likewise, this difference in life expectancy implies that, compared to high-income, low-income individuals, on average, receive a worse deal from a uniform priced pension plan because they collect the benefits of such a pension plan for a relatively shorter period. Our finding of a two-and-a-half years difference in life expectancy between low and high-income individuals is close to retirement window of two years (between ages 65 and 67) that is part of the proposed pension reforms in the Netherlands. Although the public pension benefits will be adjusted in an actuarially fair way when retiring before the newly proposed statutory retirement age of 67, through collective agreements on a sector or firm level this loss of income can be compensated via the occupational pension.[19] This may mitigate the adverse income redistribution effects that result from differential mortality. In this respect, our findings underscore the importance to allow for a retirement window and leaving it up to the individual worker when to exit the labor market within this window; a decision that will presumably depend on his or her health and life expectancy.

[19]This compensation is made possible because in low-income sectors people have a below average life expectancy and pension premiums are based on population life expectancy. We refer to Stichting van de Arbeid (2010) for details on all proposed reforms to the Dutch pension system.

Extended estimation results 6.A

Table 6.A.1: Components of individual income during retirement[a]

Male income components						
Age	Public pension %	Occupational pension %	Labor income %	Transfer income %	Capital income %	Total %
65–69	51.8	39.7	3.9	1.4	3.1	100.0
70–74	52.3	37.9	2.4	2.2	5.3	100.0
75–79	53.9	35.1	1.3	2.9	6.9	100.0
80–84	56.5	31.1	0.9	3.6	8.0	100.0
85–89	61.9	26.3	0.6	3.3	7.8	100.0
90–94	66.6	22.7	0.5	2.4	7.8	100.0
95+	67.5	25.8	1.6	0.9	4.1	100.0
Total	53.6	36.4	2.4	2.3	5.4	100.0
Female income components						
Age	Public pension %	Occupational pension %	Labor income %	Transfer income %	Capital income %	Total %
65–69	80.9	13.2	1.3	2.2	2.4	100.0
70–74	79.0	14.0	0.8	2.6	3.6	100.0
75–79	76.3	15.3	0.6	3.1	4.8	100.0
80–84	74.4	16.5	0.6	3.4	5.2	100.0
85–89	73.9	16.7	0.6	3.5	5.4	100.0
90–94	75.3	15.2	0.5	3.0	6.1	100.0
95+	75.7	15.4	0.2	3.0	5.7	100.0
Total	77.7	14.7	0.8	2.8	4.0	100.0

[a] All residents from the statutory retirement age of 65 onward receive a public retirement pension. Occupational pensions, on the other hand, are related to the individual's own employment or the employment of the spouse (in case of widowhood) before age 65. They therefore include private annuities, which may be more common among the self-employed. Labor income includes both work-related earnings and income from self-employment. Transfer income includes mainly alimony payments or receipts and rental subsidies. Capital income includes primarily interest, dividends, and income from real estate. A few income components might be considered household rather than individual income, e.g. rental subsidies. These components are recorded on an individual level by Statistics Netherlands. More details on the components of income and their definitions can be found in Knoef et al. (2009).

Table 6.A.2: Income scenarios during retirement for different household types (euros)

	Baseline situation	Difference from the baseline situation due to income differences					
	Man/Woman Median income	*Man Median +10%*	*Woman Median +10%*	*Man Low income, minimum wage*	*Woman Low income, minimum wage*	*Man High income, 2× median*	*Woman High income, 2× median*
Pension-related gross yearly salary							
Full time	29500	32450		16392		59000	
Part time	14750		16225		8196		29500
Annual income during retirement by household type							
Single person household							
Man, was employed full time before 65	20650	22715	22715	11705	11705	41300	41300
Woman, was employed part time before 65	16178		17210		11705		26503
Couple, before age 65, the man was employed full time and the woman was not employed							
Man married at age 65	16962	19027		8018		37612	
Woman married at age 65	8018	8018		8018		8018	
Household income while married	24980	27045		16035		45630	
Man becomes widowed	20650	22715		11705		41300	
Woman becomes widowed	18094	19569		11705		32844	
Couple, before age 65, the man was employed full time and the woman part time							
Man married at age 65	16962	19027	16962	8018	16962	37612	16962
Woman married at age 65	12490	12490	13522	12490	8018	12490	22815
Household income while married	29452	31517	30485	20507	24980	50102	39777
Man becomes widowed	23845	25910	24582	14900	20650	44495	31220
Woman becomes widowed	22567	24042	23599	16178	18094	37317	32892

Table 6.A.3: Estimation results for men

Mortality risk	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	Parameter estimate	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*
Constant	-3.336	0.206	-11.636	1.054	-11.398	1.037	-11.533	1.038	-10.468	0.240
Time dummy variables	included		included		included		included		included	
Age specific dummy variables[a]	included		excluded		excluded		excluded		excluded	
Age (in years)			0.132	0.016	0.128	0.015	0.130	0.015	0.112	0.003
Single	0.000		0.000		0.000		0.000		0.000	
Married	-0.923	0.142	-0.872	0.142	-0.893	0.093	-0.753	0.096	-0.407	0.099
Widowed	-0.472	0.108	-0.436	0.104	-0.425	0.103	-0.428	0.104	-0.221	0.069
$Y^I/10{,}000$	-0.218	0.042	-0.214	0.041	-0.203	0.038			-0.190	0.024
$Y^P/10{,}000$	-0.083	0.127	-0.097	0.124					-0.267	0.098
$(Y^I/10{,}000)^2$	0.008	0.002	0.008	0.002	0.009	0.002			0.009	0.001
$(Y^P/10{,}000)^2$	-0.004	0.016	-0.003	0.015					0.017	0.008
$(Y^I/10{,}000) \times (Y^P/10{,}000)$	0.028	0.016	0.028	0.016					0.020	0.012
(τ-65)	-0.034	0.036	-0.053	0.034	-0.058	0.033	-0.048	0.033		
Single x (τ-65)	0.000		0.000		0.000		0.000			
Married x (τ-65)	0.167	0.046	0.154	0.044	0.090	0.025	0.098	0.026		
Widowed x (τ-65)	0.071	0.029	0.063	0.028	0.058	0.027	0.061	0.028		
$(Y^I/10{,}000) \times (\tau\text{-}65)$	-0.162	0.115	-0.159	0.111	-0.096	0.098				
$(Y^P/10{,}000) \times (\tau\text{-}65)$	-1.025	0.543	-0.955	0.522						
$(Y^I/10{,}000)^2 \times (\tau\text{-}65)$	0.017	0.009	0.017	0.009	0.017	0.008				
$(Y^P/10{,}000)^2 \times (\tau\text{-}65)$	0.098	0.099	0.094	0.094						
$(Y^I/10{,}000) \times (Y^P/10{,}000) \times (\tau\text{-}65)$	0.064	0.092	0.065	0.088						

[a] Dummy variables are included for age 66 up to and including age 95. For ages over 95 a linear effect is included with 65 as the reference age. When these are excluded, the age variable is included.

Table 6.A.3: Estimation results for men (*continued*)

Mortality risk	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	Parameter estimate	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*
$(\tau\text{-}65)^2$	0.043	0.164	0.117	0.151	0.098	0.141	0.063	0.143		
Single x $(\tau\text{-}65)^2$	0.000		0.000		0.000		0.000			
Married x $(\tau\text{-}65)^2$	-0.725	0.258	-0.659	0.246	-0.364	0.126	-0.416	0.131		
Widowed x $(\tau\text{-}65)^2$	-0.291	0.146	-0.251	0.137	-0.230	0.136	-0.241	0.137		
$(Y^I/10{,}000) \times (\tau\text{-}65)^2$	0.075	0.063	0.078	0.060	0.084	0.053				
$(Y^P/10{,}000) \times (\tau\text{-}65)^2$	0.373	0.338	0.347	0.323						
$(Y^I/10{,}000)^2 \times (\tau\text{-}65)^2$	-0.009	0.006	-0.009	0.006	-0.011	0.005				
$(Y^P/10{,}000)^2 \times (\tau\text{-}65)^2$	-0.037	0.053	-0.035	0.051						
$(Y^I/10{,}000) \times (Y^P/10{,}000) \times (\tau\text{-}65)^2$	0.023	0.060	0.018	0.057						
$Y^H/10{,}000$							-0.196	0.038		
$(Y^H/10{,}000)^2$							0.008	0.002		
$(Y^H/10{,}000) \times (\tau\text{-}65)$							-0.165	0.103		
$(Y^H/10{,}000)^2 \times (\tau\text{-}65)$							0.017	0.008		
$(Y^H/10{,}000) \times (\tau\text{-}65)^2$							0.102	0.056		
$(Y^H/10{,}000)^2 \times (\tau\text{-}65)^2$							-0.009	0.005		
Standard deviation, random effect	0.523		0.365		0.340		0.346			
Log-likelihood value	-10872.5		-10888.9		-10902.6		-10897.1		-10914.8	
Number of observations	54617		54617		54617		54617		54617	
Number of individuals	7657		7657		7657		7657		7657	
Number of parameters	66		36		27		27		19	

Table 6.A.4: Estimation results for women

Mortality risk	**Model 6**		**Model 7**		**Model 8**		**Model 9**		**Model 10**	
	Parameter estimate	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*
Constant	-4.197	0.220	-11.337	0.891	-11.501	0.886	-11.434	0.888	-12.155	0.215
Time dummy variables	included		included		included		included		included	
Age specific dummy variables[a]	included		excluded		excluded		excluded		excluded	
Age (in years)			0.109	0.014	0.112	0.014	0.109	0.014	0.121	0.002
Single	0.000		0.000		0.000		0.000		0.000	
Married	-0.359	0.140	-0.352	0.136	-0.529	0.110	-0.153	0.110	-0.301	0.090
Widowed	0.018	0.111	0.014	0.107	0.005	0.106	0.006	0.106	-0.037	0.050
$Y^I/10{,}000$	-0.282	0.050	-0.271	0.048	-0.270	0.045			-0.186	0.030
$Y^P/10{,}000$	-0.104	0.057	-0.103	0.056					-0.091	0.040
$(Y^I/10{,}000)^2$	0.011	0.003	0.010	0.003	0.011	0.003			0.008	0.002
$(Y^P/10{,}000)^2$	0.002	0.004	0.002	0.004					0.003	0.003
$(Y^I/10{,}000) \times (Y^P/10{,}000)$	0.016	0.022	0.015	0.022					-0.003	0.016
$(\tau$-65)	0.016	0.033	0.020	0.030	0.019	0.029	0.030	0.028		
Single $\times$ (τ-65)	0.000		0.000		0.000		0.000			
Married $\times$ (τ-65)	-0.037	0.030	-0.029	0.028	-0.023	0.022	-0.037	0.022		
Widowed $\times$ (τ-65)	-0.025	0.021	-0.022	0.020	-0.020	0.020	-0.020	0.020		
$(Y^I/10{,}000) \times (\tau$-65)	0.086	0.121	0.094	0.114	0.076	0.109				
$(Y^P/10{,}000) \times (\tau$-65)	0.207	0.163	0.198	0.155						
$(Y^I/10{,}000)^2 \times (\tau$-65)	0.009	0.012	0.008	0.011	0.008	0.011				
$(Y^P/10{,}000)^2 \times (\tau$-65)	0.005	0.011	0.005	0.010						
$(Y^I/10{,}000) \times (Y^P/10{,}000) \times (\tau$-65)	-0.199	0.115	-0.187	0.109						

[a] Dummy variables are included for age 66 up to and including age 95. For ages over 95 a linear effect is included with 65 as the reference age. When these are excluded, the age variable is included.

Table 6.A.4: Estimation results for women (*continued*)

Mortality risk	**Model 6**		**Model 7**		**Model 8**		**Model 9**		**Model 10**	
	Parameter estimate	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*	*Parameter estimate*	*S.e.*
$(\tau\text{-}65)^2$	-0.184	0.117	-0.084	0.103	-0.095	0.101	-0.094	0.092		
Single × $(\tau\text{-}65)^2$	0.000		0.000		0.000		0.000			
Married × $(\tau\text{-}65)^2$	0.325	0.162	0.281	0.150	0.207	0.102	0.237	0.100		
Widowed × $(\tau\text{-}65)^2$	0.125	0.086	0.106	0.080	0.102	0.080	0.101	0.079		
$(Y^I/10{,}000) \times (\tau\text{-}65)^2$	-0.002	0.058	-0.007	0.054	0.003	0.053				
$(Y^P/10{,}000) \times (\tau\text{-}65)^2$	-0.184	0.115	-0.171	0.108						
$(Y^I/10{,}000)^2 \times (\tau\text{-}65)^2$	-0.007	0.006	-0.006	0.006	-0.006	0.006				
$(Y^P/10{,}000)^2 \times (\tau\text{-}65)^2$	0.000	0.009	0.000	0.009						
$(Y^I/10{,}000) \times (Y^P/10{,}000) \times (\tau\text{-}65)^2$	0.145	0.082	0.134	0.076						
$Y^H/10{,}000$							-0.187	0.038		
$(Y^H/10{,}000)^2$							0.006	0.002		
$(Y^H/10{,}000) \times (\tau\text{-}65)$							0.042	0.082		
$(Y^H/10{,}000)^2 \times (\tau\text{-}65)$							0.003	0.006		
$(Y^H/10{,}000) \times (\tau\text{-}65)^2$							-0.006	0.040		
$(Y^H/10{,}000)^2 \times (\tau\text{-}65)^2$							-0.002	0.003		
Standard deviation, random effect	0.507		0.233		0.233		0.232			
Log-likelihood value	-14804.9		-14823.3		-14830.4		-14831.1		-14836.5	
Number of observations	87108		87108		87108		87108		87108	
Number of individuals	11601		11601		11601		11601		11601	
Number of parameters	66		36		27		27		19	

Table 6.A.5: Simulation results when random effects and dynamic selection correction terms are excluded[a]

	Baseline situation	Difference from the baseline situation due to income differences					
	Median income	Man Median+10%	Woman Median+10%	Man Low income	Woman Low income	Man High income	Woman High income
Type of household	*Remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*	*Difference in remaining life expectancy*
Single person household							
Man	13.00 (0.52)	0.20 (0.03)		-0.90 (0.13)		1.86 (0.29)	
Woman	17.30 (0.44)		0.11 (0.02)		-0.50 (0.09)		1.09 (0.17)
Couple, before age 65, the man was employed full time and the woman was not employed							
Man	15.99 (0.44)	0.20 (0.03)		-0.92 (0.12)		1.83 (0.26)	
Woman	18.74 (0.38)	0.14 (0.03)		-0.64 (0.13)		1.26 (0.27)	
Couple, before age 65, the man was employed full time and the woman part time							
Man	16.51 (0.46)	0.19 (0.03)	0.11 (0.04)	-0.88 (0.12)	-0.53 (0.18)	1.74 (0.25)	1.09 (0.41)
Woman	19.25 (0.36)	0.13 (0.03)	0.11 (0.02)	-0.62 (0.13)	-0.51 (0.08)	1.24 (0.27)	1.09 (0.18)

[a] Estimation results are given in tables 6.A.3 (model 5) and table 6.A.4 (model 10). Tables 6.4.1 and 6.A.2 provide details on the income classifications. Standard errors are in parentheses.

6.B Monte Carlo simulations

The Monte Carlo simulations are based on the distribution function of remaining life duration at age 65 and carried out as follows. First, as a baseline, we define a group of reference individuals with the same characteristics; for instance, married men with median income. The estimates of the parameters of equation (6.3.1) together with the assumption that the error term in equation (6.3.1) follows a logistic distribution enable calculation of the probabilities that at age 65 each individual in this reference group will be deceased by the subsequent year. Next, we compare these probabilities with random drawings from the uniform distribution to simulate whether or not each individual is deceased in the subsequent year (see e.g., Law and Kelton, 1982). Finally, we simulate age by age the mortality status for each individual in the reference group up to the age of 105. We assume that next year's mortality probability is equal to one at age 105. In this way, we obtain the simulated mortality status for each individual in this group from age 65 up to age 105. We use these sequences to calculate the mean remaining life duration at 65 for this homogenous reference group of individuals. We then rerun these simulations with a change in one of the covariates (e.g., a different income level) so that, all other factors being constant, the differences between these simulation outcomes and the baseline simulation outcomes can reveal this covariate's association with remaining life expectancy at 65. For reference individuals who are married at age 65 the simulations take explicitly into account that the spouse may die and the individual becomes widowed. We perform these Monte Carlo simulations for 10,000 (identical) individuals. The standard errors for the differences from the baseline situation are based on 100 drawings from the asymptotic distribution of the parameter estimates.

Bibliography

Aaberge, R., U. Colombino, and E. Holmøy (2004): Population ageing and fiscal sustainability: An integrated micro-macro analysis of required tax changes. Discussion Paper 367, Statistics Norway. Cited on page 9.

Abbring, J. and G. van den Berg (2003): The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(3):701–710. Cited on page 52.

Akerlof, G. (1980): A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, 94:749–775. Cited on page 15.

Alessie, R., H. Camphuis, and A. Kapteyn (1990): Een simulatiemodel ter berekening van het sociale risico: eindverslag. Economics Institute Tilburg. Cited on page 113.

Apps, P. (1991): Tax reform, population ageing and the changing labour supply behaviour of married women. *Journal of Population Economics*, 4:201–216. Cited on page 9.

Attanasio, O. and C. Emmerson (2003): Mortality, health status and wealth. *Journal of the European Economic Association*, 1(4):821–850. Cited on pages 159, 160, and 163.

Attanasio, O. and H. Hoynes (2000): Differential mortality and wealth accumulation. *Journal of Human Resources*, 35(1):1–29. Cited on page 159.

Attanasio, O., H. Low, and V. Sanchez-Marcos (2008): Explaining Changes in Female Labour Supply in a Life-Cycle Model. *American Economic Review*, 98(4):1517–1552. Cited on page 15.

Attias-Donfut, C., J. Ogg, and F. Wolff (2005): Financial transfers. In B.-S. et. al., editor, *Health, Ageing and Retirement in Europe - First Results from the Survey of Health, Ageing and Retirement in Europe*, pages 179–185. Mannheim: MEA. Cited on page 85.

Baker, M., M. Stabile, and C. Deri (2004): What do self-reported, objective, measures of health measure? *Journal of Human Resources*, 39(4):1067–1093. Cited on page 40.

Baltagi, B. (1995): *Econometric analysis of panel data*. John Wiley and Sons Ltd, Chichester. Cited on page 145.

Bamia, C., A. Trichopoulou, and D. Trichopoulos (2007): Age at retirement and mortality in a general population sample: the Greek EPIC study. *American Journal of Epidemiology*, 167(5):561–569. Cited on pages 40 and 41.

Barker, D. (1995): Fetal origins of coronary heart disease. *British Medical Journal*, 311(6998):171–174. Cited on page 160.

——— (1997): Maternal nutrition, fetal nutrition, and diseases in later life. *Nutrition*, 13(9):807–813. Cited on page 160.

Bell, D., A. Dowes, and A. Heitmueller (2006): Did the introduction of free personal care in Scotland result in a reduction of informal care? Paper presented at IZA conference on the Well-Being of the Elderly. Cited on page 66.

Benati, L. (2001): Some empirical evidence of the discouraged worker effect. *Economics Letters*, 70:387–395. Cited on page 16.

Bernheim, B., A. Shleifer, and L. Summers (1985): The strategic bequest motive. *Journal of Political Economy*, 93:1045–1076. Cited on page 68.

Blakely, T., I. Kawachi, J. Atkinson, and J. Fawcett (2004): Income and mortality: the shape of the association and confounding New Zealand Census-Mortality Study, 1981–1999. *International Journal of Epidemiology*, 33(8):874–883. Cited on page 159.

Blundell, R. and T. MaCurdy (1999): Labour supply: A review of alternative approaches. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, vol. 3A*, pages 1559–1695. North-Holland. Cited on page 12.

Bolin, K., B. Lindgren, and P. Lundborg (2008a): Informal and formal care among single-living elderly in Europe. *Health Economics*, 17(3):393–409. Cited on pages 66 and 77.

——— (2008b): Your next of kin or your own career? Caring and working among the 50+ of Europe. *Journal of Health Economics*, 27(3):718–738. Cited on pages 68, 71, 74, 77, and 87.

Bonenkamp, J. (2009): Measuring lifetime redistribution in Dutch occupational pensions. *De Economist*, 157(1):49–77. Cited on page 160.

Bonsang, E. (2007): How do middle-aged children allocate time and money transfers to their older parents in Europe? *Empirica*, 34:171–188. Cited on pages 66, 77, and 85.

——— (2009): Does informal care from children to their elderly parents substitute for formal care in Europe. *Journal of Health Economics*, 28:143–154. Cited on pages 66 and 77.

Bonsang, E., S. Adam, and S. Perelman (2010): Does retirement affect cognitive functioning. Working paper, Maastricht University, RM/10/005. Cited on page 42.

Bound, J. (1991): Self-reported versus objective measures of health in retirement models. *Journal of Human Resources*, 26(1):106–138. Cited on page 40.

Bound, J., M. Schoenbaum, T. Stinebrickner, and T. Waidmann (1999): The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, 6(2):179–202. Cited on page 40.

Bourguignon, F. (1979): Decomposable income inequality measures. *Econometrica*, 47:901–920. Cited on page 158.

Bovenberg, A. and H. ter Rele (2000): Generational accounts for the Netherlands: an update. *International tax and public finance*, 7(4):411–430. Cited on page 109.

Brockman, H., R. Muller, and U. Helmert (2009): Time to retire - Time to die? A prospective cohort study of the effects of early retirement on long-term survival. *Social Science & Medicine*, 69:160–164. Cited on pages 40 and 41.

Brown, J. (2002): Differential mortality and the value of individual account retirement annuities. In M. Feldstein and J. Liebman, editors, *The distributional aspects of social security and social security reform*, chapter 10, pages 401–445. University of Chicago Press. Http://www.nber.org/chapters/c9756. Cited on page 160.

Buhmann, B., L. Rainwater, G. Schmaus, and T. Smeeding (2005): Equivalence scales, well-being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg income study (LIS) database. *Review of Income and Wealth*, 34(2):115–142. Cited on page 114.

Butricia, B., H. Iams, J. Moore, and M. Waid (2001): Methods in Modeling Income in the Near Term (MINT I). ORES working paper n.91. Cited on pages 123 and 125.

Byrne, D., M. Goeree, B. Hiedemann, and S. Stern (2009): Formal home health care, informal care, and family decision making. *International Economic Review*, 50(4):1205–1242. Cited on pages 64, 68, and 70.

Callegaro, L. and G. Pasini (2008): Social Interaction effects in an intergenerational model of informal care giving. Netspar Discussion Paper 2008-023. Cited on pages 64 and 68.

Cameron, A. and P. Trivedi (2005): *Microeconometrics: methods and applications*. Cambridge University Press. Cited on pages 51, 145, and 172.

Cameron, S. and J. Heckman (1998): Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males. *Journal of Political Economy*, 106(2):262–333. Cited on pages 162 and 170.

Caminada, C. and K. Goudswaard (2001): International trends in income inequality and social policy. *International Tax and Public Finance*, 8:395–415. Cited on page 112.

——— (2003): *Verdeelde zekerheid. De verdeling van baten en lasten van sociale zekerheid en pensioen*. E.M. Meijers Instituut, Sdu Uitgevers. Cited on page 112.

Campolieti, M. (2002): Disability and the labor force participation of older men in Canada. *Labour Economics*, 9(3):405–432. Cited on page 40.

Case, A., A. Fertig, and C. Paxson (2005): The lasting impact of childhood health and circumstances. *Journal of Health Economics*, 24(2):365–389. Cited on pages 160 and 181.

Case, A., D. Lubotsky, and C. Paxson (2002): Economic status and health in childhood: the origins of the gradient. *American Economic Review*, 92(5):1308–1334. Cited on pages 160 and 181.

CBS (2009a): Documentatierapport Inkomenspanel onderzoek (IPO). Centraal Bureau voor de Statistiek, Voorburg. Cited on pages 43, 113, and 163.

——— (2009b): Documentatierapport Doodsoorzaken (DO). Centraal Bureau voor de Statistiek, Voorburg. Cited on pages 43 and 163.

——— (2010): Documentatierapport Gemeentelijke Basisadministratie (GBA) 1995-2010V1. Centraal Bureau voor de Statistiek, Voorburg. Cited on page 113.

Charles, K. and P. Sevak (2005): Can family caregiving substitute for nursing home care. *Journal of Health economics*, 24(6):1174–1190. Cited on page 71.

Checkovich, T. and S. Stern (2002): Shared caregiving responsibilities of adult siblings with elderly parents. *The Journal of Human Resources*, 37(3):441–478. Cited on page 64.

Coe, N. and M. Lindeboom (2008): Does retirement kill you? Evidence from early retirement windows. IZA discussion paper series, No. 3817. Cited on page 42.

Coe, N. and G. Zamarro (2008): Retirement effects on health in Europe. RAND Labor and Population working paper series, wr-588. Cited on page 42.

Coleman, M. and J. Pencavel (1993): Trends in market work behavior of women since 1940. *Industrial and Labor Relations Review*, 46:653–676. Cited on page 14.

Colombino, U. and B. de Stavola (1985): A model of female labor supply in Italy using cohort data. *Journal of Labor Economics*, 3:S275–S292. Cited on page 14.

Connidis, I. and L. Campbell (1995): Closeness, confiding, and contact among siblings in middle and late adulthood. *Journal of Family Issues*, 16(6):722–745. Cited on page 100.

Cörvers, F. and B. Golsteyn (2003): Changes in women's willingness to work in a tightening labour market: the impact of preferences, wages and individual characteristics. ROA Research Memorandum 2003/5E. Cited on page 13.

Cournède, B. and F. Gonand (2006): Restoring fiscal sustainability in the Euro Area: raise taxes or curb spending? Economics Department Working Paper No. 520, OECD. Cited on page 10.

Cowell, F. (1980): On the structure of additive inequality measures. *Review of Economic Studies*, 47:521–531. Cited on page 158.

CPB (2005): Centraal Economisch Plan. Netherlands Bureau for Economic Policy Analysis. Chapter 6.3: Stelselwijziging VUT, prepensioen en levensloop. Cited on page 39.

——— (2006): Het groeipotentieel van de Nederlandse economie tot 2011. CPB Document 117. Cited on page 30.

——— (2009): Houdbaarheidseffect verhoging AOW leeftijd. CPB note June, 19, Netherlands Bureau for Economic Policy Analysis, The Hague. Cited on page 39.

——— (2010): Analyse voorstel verhoging AOW-leeftijd. Netherlands Bureau for Economic Policy Analysis, CPB notitie January 13. Cited on page 39.

Creedy, J. and A. Duncan (2002): Behavioural microsimulation with labour supply responses. *Journal of economic surveys*, 16(1). Cited on page 112.

Crespo, L. and P. Mira (2010): Caregiving of elderly parents and employment status of European mature women. Working paper presented at the Econometric Society World Congress. Cited on page 77.

Cutler, D., A. Deaton, and A. Lleras-Muney (2006): The determinants of mortality. *Journal of Economic Perspectives*, 20(3):97–120. Cited on page 53.

Darby, J., R. Hart, and M. Vecchi (2001): Labour force participation and the business cycle: A comparative analysis of France, Japan, Sweden, and the United States. *Japan and the World Economy*, 13:113–133. Cited on page 16.

De Jong, A. (2005): Bevolkingsprognose 2004-2050: maximaal 17 miljoen inwoners. *Bevolkingstrends*, 53(1):12–18. Cited on page 30.

De Kam, F. and F. Nijpels (1995): *Tijdbom: vergrijzing dreigt de AOW, onze pensioenen en de ouderenzorg onbetaalbaar te maken*. Uitgeverij Contact, Amsterdam. Cited on page 120.

De Meijer, C., M. Koopmanschap, X. Koolman, and E. Van Doorslaer (2009): The role of disability in explaining long-term care utilization. *Medical Care*, 47(11):1156–1163. Cited on page 87.

De Vos, K. (2007): 'Remote Access' final report. *Netspar pilot study*. Cited on page 112.

De Vos, K., A. Kalwij, and A. Kapteyn (2010): Disability insurance and labor market exit routes of older workers in the Netherlands. Working paper, Tilburg University. Cited on pages 39, 46, 60, and 61.

Deaton, A. and C. Paxson (1994): Saving, growth and aging in Taiwan. In D. Wise, editor, *Studies in the Economics of Aging*, chapter 5, pages 331–362. National Bureau of Economic Research. Cited on page 127.

Deelen, A. (1995): De ontwikkeling van de inkomensverdeling in Nederland op lange termijn. Onderzoeksmemorandum 121, CPB. Cited on page 120.

Dekkers, G., H. Busslei, M. Cozzolino, R. Desmet, J. Geyer, D. Hoffmann, M. Raitano, V. Steiner, P. Tanda, S. Tedeschi, and F. Verschueren (2008): A classification and overview of micro simulation models, and the choices made in MIDAS. In *What are the consequenses of the AWG projections for the adequacy of social security pensions*, Report of the Work Package 4 of the AIM project. Cited on page 123.

Den Butter, F. (2010): Een waardevaste AOW is generatiebewust. *Economisch statistische berichten*, 95(4579):102–105. Cited on page 109.

Dessens, J. and W. Jansen (1997): Future trends in household income inequality in the Netherlands. *The Netherlands journal of social sciences*, 33(2):113–129. Cited on page 110.

Drukker, D. and R. Gates (2006): Generating Halton sequences using Mata. *The Stata Journal*, 6(2):214–228. Cited on page 74.

Duleep, H. (1986): Measuring income's effect on adult mortality using longitudinal administrative record data. *Journal of Human Resources*, 21(2):238–251. Cited on pages 49 and 159.

Dykstra, P. (1993): The differential availability of relationships and the provision and effectiveness of support to older adults. *Journal of Social and Personal Relationships*, 10:355–370. Cited on page 67.

Eggebeen, D. and D. Hogan (1990): Giving between generations in American families. *Human Nature*, 1:211–232. Cited on page 67.

Eliason, M. and D. Storrie (2009): Does job loss shorten life? *Journal of Human Resources*, 44(2):277–302. Cited on page 42.

Emmerson, C., H. Reed, and A. Shephard (2004): An assessment of PENSIM2. IFS working paper WP04/21, London: Institute for fiscal studies. Cited on page 123.

Engers, M. and S. Stern (2002): Long-term care and family bargaining. *International Economic Review*, 43(1):73–114. Cited on page 64.

Ettner, S. (1995): The impact of 'parent care' on female labor supply decisions. *Demography*, 32(1):63–80. Cited on page 68.

——— (1996): The opportunity cost of elder care. *The Journal of Human Resources*, 31(1):189–205. Cited on page 68.

European Commission (2009): Sustainability report. European Economy 9, Economic and Financial Affairs. Cited on page 2.

Euwals, R., R. de Mooij, and D. van Vuuren (2009): Rethinking Retirement. CPB Netherlands Bureau for Economic Policy Analysis. Cited on page 39.

Euwals, R., M. Knoef, and D. van Vuuren (2011): The trend in female labour force participation: what can be expected for the future? *Empirical Economics*, 40(3):729–753. Cited on pages 9 and 117.

Euwals, R., D. van Vuuren, A. den Ouden, and J. Rijn (2006): Arbeidsaanbod naar sociaaldemografische kenmerken. CPB Memorandum 171. Cited on page 18.

Evers, M., R. de Mooij, and D. van Vuuren (2008): The wage elasticity of labour supply: A synthesis of empirical estimates. *De Economist*, 156(1):25–43. Cited on page 91.

Fernandez, R. (2007): Culture as learning: the evolution of female labour force participation over a century. CEPR Discussion Paper 6451. Cited on page 35.

Fernandez, R., A. Fogli, and C. Olivetti (2004): Preference formation and the rise of women's labour force participation: Evidence from WWII. NBER Working Paper 10589. Cited on page 15.

Fitzenberger, B., R. Schnabel, and G. Wunderlich (2004): The gender gap in labor market participation and employment: A cohort analysis for West Germany. *Journal of Population Economics*, 17(1):83–116. Cited on pages 13 and 20.

Fitzenberger, B. and G. Wunderlich (2004): The changing life cycle pattern in female employment: a comparison of Germany and the UK. *Scottish Journal of Political Economy*, 51:302–328. Cited on page 13.

Flood, L., A. Klevmarken, and A. Mitrut (2008): The income of the Swedish baby boomers. In A. Klevmarken and B. Lindgren, editors, *Simulating an ageing population. A microsimulation approach to Sweden*, number 285 in Contributions to Economic Analysis, chapter 8, pages 249–292. Emerald Group Publishing Limited. Cited on page 123.

Fogli, A. and L. Veldkamp (2007): Nature or nurture? Learning and female labour force dynamics. CEPR Discussion Paper 6324. Cited on page 15.

Fontaine, R., A. Gramain, and J. Wittwer (2009): Providing care for an elderly parent: interactions among siblings. *Health economics*, 18(9):1011–1029. Cited on page 64.

Frankenberg, E., L. Lillard, and R. Willis (2002): Patterns of intergenerational transfers in southeast Asia. *Journal of Marriage and the Family*, 64:627–641. Cited on page 66.

Geyer, J. and V. Steiner (2010): Public pensions, changing employment patterns, and the impact of pension reforms across birth cohorts: a microsimulation analysis for Germany. IZA discussion paper, no. 4815. Cited on page 123.

Gjesdal, S., J. Maeland, J. Hagberg, and K. Alexanderson (2007): Socioeconomic inequalities and mortality among disability pensioners in Norway - a population-based cohort study. *Norsk Epidemiologi*, 17(1):29–35. Cited on page 41.

Goldin, C. (2004): The long road to the fast track: career and family. *The Annals of the American Academy of Political and Social Science*, 596:20–35. Cited on pages 12 and 15.

——— (2006): The 'Quiet Revolution' that transformed women's employment, education, and family. *American Economic Review, Papers and Proceedings*, 96:1–21. Cited on pages 12 and 15.

Goldin, C. and L. Katz (2002): The power of the pill: oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 10:730–770. Cited on pages 11, 15, and 35.

Gottschalk, P. and T. Smeeding (2000): Empirical evidence on income inequality in industrialized countries. In A. Atkinson and F. Bourguignon, editors, *Handbook of income distribution*, volume 1, chapter 5, pages 262–307. Elsevier, New York. Cited on pages 112 and 116.

Gregg, P. (1994): Out for the count: a social scientist's analysis of unemployment statistics in the UK. *Journal of the Royal Statistical Society Series A*, 157:253–270. Cited on page 16.

Groot, W. and H. Pott-Buter (1993): Why married women's labor supply in the Netherlands has increased? *De Economist*, 141:238–255. Cited on page 13.

Grossman, M. (1972): On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255. Cited on page 161.

——— (2000): The human capital model. In A. Culyer and J. Newhouse, editors, *Handbook of Health Economics*, volume 1A, pages 347–408. Elsevier Science, North Holland, New York, and Oxford. Cited on pages 41, 47, and 161.

Grundy, E. (2005): Reciprocity in relationships: Socio-economic and health influences on intergenerational exchanges between third age parents and their adult children in Great Britain. *British Journal of Sociology*, 56:233–255. Cited on page 67.

Grundy, E. and G. Holt (2001): The socioeconomic status of older adults: how should we measure it in studies of health inequalities? *Journal of Epidemiology and Community Health*, 55:895–904. Cited on page 49.

Hall, R. (1971): The Measurement of Quality Change from Vintage Price Data. In Z. Griliches, editor, *Price indexes and Quality Change*, chapter 8, pages 240–271. Harvard University Press. Cited on page 20.

Harding, A. (2007): Challenges and Opportunities of Dynamic Microsimulation Modelling. Plenary paper presented at the first General Conference of the International Microsimulation Association, Vienna, 21 August 2007. Cited on page 123.

Hári, N. (2007): *Modeling mortality: Empirical studies on the effect of mortality on annuity markets*. Ph.D. thesis, Tilburg University. Cited on page 160.

Hartog, J. and J. Theeuwes (1985): The emergence of the working wife in Holland. *Journal of Labor Economics*, 3:S235–S255. Cited on pages 12 and 14.

Haveman, R., K. Holden, A. Romanov, and B. Wolfe (2007): Assessing the maintenance of savings sufficiency over the first decade of retirement. *International tax and public finance*, 14:481–502. Cited on page 125.

Heckman, J. (1979): Sample selection bias as a specification error. *Econometrica*, 47(1):153–161. Cited on page 74.

Heckman, J. and B. Honoré (1989): The identifiability of the competing risks model. *Biometrika*, 76(2):325–330. Cited on page 52.

Henkens, K., Y. Grift, and J. Siegers (2002): Changes in female labour supply in the Netherlands 1989-1998: The case of married and cohabiting women. *European Journal of Population*, 18:39–57. Cited on page 13.

Hiedemann, B. and S. Stern (1999): Strategic play among family members when making long-term care decisions. *Journal of economic behavior and organization*, 40:29–57. Cited on page 64.

Hoerger, T., G. Picone, and F. Sloan (1996): Public subsidies, private provision of care and living arrangements of the elderly. *Review of Economics and Statistics*, 78(3):428–440. Cited on page 64.

Hogan, D. and D. Eggebeen (1995): Sources of emergency help and routine assistance in old-age. *Social Forces*, 73:917–936. Cited on page 65.

Huisman, M., A. Kunst, O. Andersen, M. Bopp, J.-K. Borgan, C. Borrell, G. Costa, P. Deboosere, G. Desplanques, A. Donkin, S. Gadeyne, C. Minder, E. Regidor, T. Spadea, T. Valkonen, and J. Mackenbach (2004): Socioeconomic inequalities in mortality among elderly people in 11 European populations. *Journal of Epidemiology and Community Health*, 58:468–475. Cited on pages 49 and 159.

Huisman, M., A. Kunst, M. Bopp, J.-K. Borgan, C. Borrell, G. Costa, P. Deboosere, S. Gadeyne, M. Glickman, C. Marinacci, C. Minder, E. Regidor, T. Valkonen, and J. Mackenbach (2005): Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *Lancet*, 365:493–500. Cited on pages 47, 49, and 160.

Huisman, M. and J. Mackenbach (2007): Intelligence and the explanation for socio-economic inequalities in health. *Nederlands Tijdschrift voor Geneeskunde*, 151(19):1076–1082. Cited on page 160.

Hupfeld, S. (2009): Rich and healthy - better than poor and sick?: An empirical analysis of income, health, and the duration of the pension benefit spell. *Journal of Health Economics*, 28:427–443. Cited on pages 159 and 162.

Hurd, M. (1990): The Joint Retirement Decision of Husbands and Wives. In D. Wise, editor, *Issues in the Economics of Aging*, pages 231–254. The University of Chicago Press. Cited on page 26.

Hurd, M., D. McFadden, and A. Merrill (2001): Predictors of mortality among elderly. In D. Wise, editor, *Themes in the Economics of Aging*, chapter 5, pages 171–198. University of Chicago Press. Cited on page 159.

Israel, R., H. Rosenberg, and L. Curtin (1986): Analytical potential for multiple-cause-of-death data. *American Journal of Epidemiology*, 124(2):161–179. Cited on page 52.

Iversen, L., O. Andersen, P. Andersen, K. Christoffersen, and N. Keiding (1987): Unemployment and mortality in Denmark. *British Medical Journal*, 295:879–884. Cited on page 40.

Janssen, F., J. Mackenbach, and A. Kunst (2004): Trends in old-age mortality in seven European countries, 1950-1999. *Journal of Clinical Epidemiology*, 57(2):203–216. Cited on page 53.

Jongen, E. (2010): Child care subsidies revisited. CPB Document No. 200. Cited on page 32.

Joshi, H., R. Layard, and S. Owen (1985): Why are more women working in Britain? *Journal of Labor Economics*, 3:S147–S175. Cited on page 14.

Kalmijn, M. (2006): Educational inequality and family relationships: Influences on contact and proximity. *European Sociological Review*, 22:1–16. Cited on pages 71 and 88.

——— (2007): Gender differences in the effects of divorce, widowhood, and remarriage on intergenerational support: Does marriage protect fathers? *Social Forces*, 85(3):1079–1104. Cited on page 68.

——— (2010): Verklaring van intergenerationele solidariteit: een overzicht van concurrerende theorieën en hun ondezoeksbevindingen. *Mens en Maatschappij*, 85(1):70–98. Cited on page 72.

Kalmijn, M. and R. Alessie (2008): Life course changes in income: an exploration of age- and stage effects in a 15-year panel in the Netherlands. Netspar Panel Paper 10. Cited on pages 114, 127, 132, and 133.

Kalmijn, M. and P. Dykstra (2006): Differentials in face-to-face contact between parents and their grown-up children. In P. Dykstra, M. Kalmijn, T. Knijn, A. Komter, A. Liefbroer, and C. Mulder, editors, *Family solidarity in the Netherlands*, pages 63–88. Amsterdam: Dutch University Press. Cited on page 67.

Kalmijn, M. and C. Saraceno (2008): A comparative perspective on intergenerational support: Responsiveness to parental needs in individualistic and familialistic countries. *European Societies*, 10(3):479–508. Cited on pages 67 and 68.

Kalwij, A., R. Alessie, and M. Knoef (2009): The association between individual income and remaining life expectancy at the age of 65 in the Netherlands. Netspar Discussion Paper 09/2009-033, www.netspar.nl. Cited on pages 47, 49, 128, and 159.

——— (2010): Pathways to retirement, income, and cause specific mortality risks in the Netherlands. Netspar Discussion Paper 08/2010-033, www.netspar.nl. Cited on page 39.

Kapteyn, A., R. Alessie, and A. Lusardi (2005): Explaining the wealth holdings of different cohorts: productivity growth and social security. *European Economic Review*, 49:1361–1391. Cited on pages 20 and 125.

Kapteyn, A., K. de Vos, and A. Kalwij (2010): Early retirement and employment of the young in the Netherlands. In J. Gruber and D. Wise, editors, *Social Security Programs and Retirement around the World: The relationship to youth employment*, chapter 8. University of Chicago Press. Cited on page 119.

Karlsson, N., J. Carstensen, S. Gjesdal, and K. Alexanderson (2007): Mortality in relation to disability pension: findings from a 12-year prospective population-based cohort study in Sweden. *Scandinavian Journal of Public Health*, 35:341–347. Cited on page 41.

Katsahian, S., M. Resche-Reigon, S. Chevret, and R. Porcher (2006): Analysing multicenter competing risks data with a mixed proportional hazard model for the subdistribution. *Statistics in Medicine*, 25:4267–4278. Cited on page 42.

Kerkhofs, M., M. Lindeboom, and J. Theeuwes (1999): Retirement, financial incentives and health. *Labour Economics*, 6(2):203–227. Cited on page 40.

Klein Ikkink, K., T. Van Tilburg, and K. Knipscheer (1999): Perceived instrumental support exchanges in relationships between elderly parents and their adult children: Normative and structural explanations. *Journal of Marriage and the Family*, 4:831–844. Cited on pages 67 and 85.

Knoef, M., R. Alessie, and A. Kalwij (2009): Changes in the income distribution of the Dutch elderly between 1989 and 2020: a microsimulation. Netspar Discussion Paper 09/2009-030, www.netspar.nl. Cited on pages 47, 109, and 165.

Knoef, M. and P. Kooreman (2011): The effects of cooperation; A structural model of siblings' caregiving interactions. IZA Discussion paper No. 5733. Cited on page 63.

Knoef, M. and K. de Vos (2009b): Representativeness in online panels: how far can we reach. Paper presented at the MESS workshop 2008, www.lissdata.nl. Cited on pages 113 and 123.

Kohli, M. (1999): Private and public transfers between generations: Linking the family and the state. *European Societies*, 1:81–104. Cited on pages 66 and 68.

Kohli, M. and H. Künemund (2003): Intergenerational transfers in the family: what motives giving. In V. Bengtson and A. Lowenstein, editors, *Families, aging and social supports: International perspectives*, pages 123–142. Hawthorne, New York: Aldine De Gruyter. Cited on page 72.

Kotlikoff, L. and J. Morris (1990): Why don't the elderly live with their children? A new look. In D. Wise, editor, *Issues in the Economics of Aging*, chapter 5, pages 149–172. University of Chicago Press, Chicago. Cited on page 64.

Kremer, M. (2005): *How welfare states care: culture, gender and citizenship in Europe.* Ph.D. thesis, Utrecht University. Cited on page 24.

Kuhn, A., J.-P. Wuellrich, and J. Zweimüller (2010): Fatal attraction? Access to early retirement and mortality. Working paper, University of Zürich. Cited on page 42.

Kunst, A., V. Bos, O. Andersen, M. Cardano, G. Costa, S. Harding, O. Hemström, R. Layte, E. Regidor, A. Reid, P. Santana, T. Valkonen, and J. Mackenbach (2004): Monitoring of trends in socioeconomic inequalities in mortality: experiences from a European project. *Demographic Research*, Special Collection 2:229–254. Cited on page 159.

Kunst, A. and J. Mackenbach (1994): The size of mortality differences associated with educational level in nine industrialized countries. *American Journal of Public Health*, 84:932–937. Cited on page 161.

Law, A. and W. Kelton (1982): *Simulation Modelling and Analysis*. New York: McGraw-Hill. Cited on pages 128 and 190.

Layard, R., M. Barton, and A. Zabalza (1980): Married women's participation and hours. *Economica*, 47:51–72. Cited on page 14.

Lee, Y., W. Parish, and R. Willis (1994): Sons, daughters, and intergenerational support in Taiwan. *American Journal of Sociology*, 99:1010–1041. Cited on page 66.

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(1):144–168. Cited on pages 49 and 159.

Lindeboom, M. and M. Kerkhofs (2009): Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work. *Journal of Applied Econometrics*, 24(6):1024–1046. Cited on pages 40 and 41.

Litwin, H. (2007): Does early retirement lead to longer life? *Ageing & Society*, 27:739–754. Cited on page 40.

Macintyre, S. (1997): The Black Report and beyond: what are the issues? *Social Science and Medicine*, 44(6):723–745. Cited on pages 159 and 160.

Mackenbach, J., A. Kunst, H. Lautenbach, F. Bijlsma, and Y. Oei (1995): Competing causes of death: an analysis using multiple-cause-of-death data from the Netherlands. *Journal of Epidemiology*, 141(5):466–475. Cited on pages 48 and 51.

——— (1999): Gains in life expectancy after elimination of major causes of death: revised estimates taking into account the effect of competing causes. *Journal of Epidemiology and Community Health*, 53:32–37. Cited on page 52.

Mackenbach, J., A. Kunst, H. Lautenbach, Y. Oei, and F. Bijlsma (1997): Competing causes of death: a death certificate study. *Journal of Clinical Epidemiology*, 50(10):1067–1077. Cited on page 48.

Maloney, T. (1991): Unobserved variables and the elusive added worker effect. *Economica*, 58:173–187. Cited on page 16.

——— (1997): Employment constraints and the labor supply of married women: A reexamination of the added worker effect. *Journal of Human Resources*, 22:51–61. Cited on page 16.

Marmot, M., G. Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, and A. Feeny (1991): Health inequalities among British civil servants: The Whitehall II study. *Lancet*, pages 1387–1393. Cited on pages 49, 53, 159, and 160.

Martikainen, P., P. Mäkelä, S. Koskinen, and T. Valkonen (2001): Income differences in mortality: a register-based follow-up study of three million men and women. *International Journal of Epidemiology*, 30:1397–1405. Cited on pages 159, 161, and 181.

Martini, A. and U. Trivellato (1997): The role of survey data in microsimulation models for social policy analysis. *Labour*, 11(1):83–112. Cited on page 123.

Mason, K., W. Mason, H. Winsborough, and W. Poole (1973): Some methodological issues in cohort analysis of archival data. *American Sociological Review*, 38:242–258. Cited on page 21.

McDonough, P., D. Williams, and J. House (1999): Gender and the socioeconomic gradient in mortality. *Journal of Health and Social Behavior*, 40(1):17–31. Cited on page 162.

McGarry, K. (1999): Inter vivos transfers and intended bequests. *Journal of Public Economics*, 73(3):321–351. Cited on page 77.

McKelvey, R. and T. Palfrey (1995): Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior*, 10:6–38. Cited on page 94.

——— (1998): Quantal Response Equilibria for Extensive Form Games. *Experimental Economics*, 1:9–41. Cited on page 94.

Menchik, P. (1993): Economic status as a determinant of mortality among black and white older men: Does poverty kill? *Population Studies*, 47:427–436. Cited on pages 49, 159, and 160.

Merz (1991): Microsimulation - a survey of principles, developments and applications. *International Journal of Forecasting*, 7:77–104. Cited on page 110.

Nakamura, A. and M. Nakamura (1994): Predicting female labor supply: effects of children and recent work experience. *Journal of Human Resources*, 29:304–327. Cited on page 24.

Nelissen, J. (1999): Mortality differences related to socioeconomic status and the progressivity of old-age pensions and health insurance: the Netherlands. *European Journal of Population*, 15:77–97. Cited on pages 160 and 161.

Norton, C. and C. Van Houtven (2006): Inter-vivos transfers and exchange. *Southern Economic Journal*, 73(1):157–172. Cited on page 77.

O'Donoghue, C. (2001): Dynamic Microsimulation: A Methodological Survey. *Brazilian Electronic Journal of Economics*, 4(2). Cited on pages 110 and 124.

Osler, M., E. Prescott, M. Grønbæk, U. Christensen, P. Due, and G. Engholm (2002): Income inequality, individual income, and mortality in Danish adults: analysis of pooled data from two cohort studies. *British Medical Journal*, 324:13–16. Cited on pages 128 and 159.

Palme, M. and S. Sandgren (2008): Parental income, lifetime income, and mortality. *Journal of the European Economic Association*, 6(4):890–911. Cited on page 159.

Panis, C. and L. Lillard (1999): Near term model development part II, Final report. Santa Monica: Rand Corp. Cited on page 123.

Pencavel, J. (1998): The market work behavior and wages of women: 1975-1994. *Journal of Human Resources*, 33:771–804. Cited on page 14.

Perozek, M. (1998): A reexamination of the strategic bequest motive. *Journal of Political Economy*, 106(2):423–445. Cited on page 68.

Pezzin, L. and B. Schone (1997): The allocation of resources in intergenerational households: Adult children and their elderly parents. *American Economic Review*, 87(2):440–464. Cited on page 68.

——— (1999): Intergenerational household formation, female labor supply and informal caregiving: A bargaining approach. *The Journal of Human Resources*, 34(3):475–503. Cited on pages 64 and 68.

Pfau-Effinger, B. (1993): Modernisation, culture and part-time employment: The example of Finland and Germany. *Work, Society and Employment*, 7:383–410. Cited on page 12.

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(4):525–536. Cited on page 41.

Prieto-Rodriguez, J. and C. Podriguez-Gutierrez (2000): The added worker effect in the Spanish case. *Applied Economics*, 31:1917–1925. Cited on page 16.

Reher, D. (1998): Family ties in Western Europe. *Population and development review*, 24:203–234. Cited on page 88.

Ridder, G. (1984): The distribution of single-spell duration data. In G. Neumann and N. Westergard-Nielsen, editors, *Studies in Labor Market Analysis*. Springer-Verlag, Berlin. Cited on page 170.

Roeger, W. (2002): Ageing and pensions in Europe. mimeo, European Commission, Directorate General for Economic and Financial Affairs. Cited on page 10.

Rohwedder, S. and R. Willis (2010): Mental retirement. *Journal of Economic Perspectives*, 24(1):119–138. Cited on page 42.

Rother, P., M. Catenaro, and G. Schwab (2002): Ageing and pensions in the Euro area; survey and projection results. mimeo, European Central Bank. Cited on page 10.

Schirle, T. (2008): Why have the labor participation rates of older men increased since the mid-1990s? *Journal of Labor Economics*, 26(4):549–594. Cited on page 26.

SCP (2003): Inkomen verdeeld: Trends in ongelijkheid, herverdeling en dynamiek. Netherlands Institute for Social Research, number 2003/3. Cited on page 112.

——— (2006): Rapportage ouderen 2006; veranderingen in de leefsituatie en levensloop. Netherlands Institute for Social Research, number 2006/12. Cited on page 119.

SCP/CBS (2006): Emancipatiemonitor 2006. Netherlands Institute for Social Research, number 2006/22. Cited on pages 11, 25, and 35.

Shorrocks, A. (1980): The class of additively decomposable inequality measures. *Econometrica*, 48:613–625. Cited on page 158.

Siermann, C., P. van Teeffelen, and L. Urlings (2004): Equivalentiefactoren 1995-2000: Methode en belangrijkste uitkomsten. Statistics Netherlands, Sociaal-economische trends, 3, p. 63-66. Cited on pages 47 and 114.

Silverstein, M. (1995): Stability and change in temporal distance between the elderly and their children. *Demography*, 32:29–46. Cited on page 67.

Silverstein, M. and V. Bengtson (1997): Intergenerational solidarity and the structure of adult child-parent relationships in American families. *American Journal of Sociology*, 103:429–460. Cited on page 65.

Simonovits, A. (2006): Optimal design of pension rule with flexible retirement: The two-type case. *Journal of Economics*, 89(3):197–222. Cited on page 160.

Sloan, F., G. Picone, and J. Hoerger (1997): The supply of children's time to disabled elderly parents. *Economic Inquiry*, 35(2):295–308. Cited on page 68.

Smith, J. (1999): Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives*, 13(2):145–166. Cited on pages 49, 53, 159, and 160.

Smith, J. and M. Ward (1985): Time-Series growth in the female labor force. *Journal of Labor Economics*, 3:S59–S90. Cited on pages 14 and 20.

Smith, K., M. Favreault, C. Ratcliffe, B. Butricia, and E. Toder (2007): Modeling income in the near term 5. The Urban Institute, Washington D.C. Cited on page 123.

Snyder, S. and W. Evans (2006): The effect of income on mortality: evidence from the social security notch. *The Review of Economics and Statistics*, 88(3):482–495. Cited on pages 42 and 159.

Soede, A., J. Vrooman, P. Ferrasi, and G. Segre (2004): Unequal welfare states: distributive consequences of population ageing in six European countries. Netherlands Institute for Social Research (SCP) and Center for research on pensions and welfare policies (CeRP). Cited on page 124.

Spitze, G. and J. Logan (1989): Gender differences in family support: Is there a payoff? *The Gerontologist*, 29:108–113. Cited on page 67.

——— (1991): Sibling structure and intergenerational relations. *Journal of Marriage and the Family*, 53:871–884. Cited on pages 68 and 93.

Stephens, M. (2002): Worker displacement and the added worker effect. *Journal of Labor Economics*, 20(3):504–537. Cited on page 16.

Stern, S. (1989): Measuring the effect of disability on labor force participation. *Journal of Human Resources*, 24(3):361–395. Cited on page 40.

Stichting van de Arbeid (2010): Pension Accord Spring 2010 (Pensioen Akkoord). www.stvda.nl. Cited on pages 160 and 182.

Stimulansz (2009): *De Grote Almanak*. Utrecht, July. Cited on page 39.

Sullivan, D. and T. Von Wachter (2009): Job displacement and mortality: an analysis using administrative data. *Quarterly Journal of Economics*, 124(3):1265–1306. Cited on pages 42 and 159.

Sundström, M. (1991): Part-time work in Sweden: Trends and equality effects. *Journal of Economic Issues*, 25:167–187. Cited on page 12.

SZW (2006): De toekomstige inkomenspositie van ouderen. Ministry of Social Affairs and Employment. Cited on page 110.

——— (2009): Onderzoek effecten afschaffing AOW-partnertoeslag. Ministry of Social Affairs and Employment. Cited on page 128.

Tachibanaki, T. and K. Sakurai (1991): Labour supply and unemployment in Japan. *European Economic Review*, 35:1575–1587. Cited on page 16.

Toder, E., C. Uccello, J. O'Hare, M. Favreault, C. Ratcliff, K. Smith, G. Burtless, and B. Bosworth (1999): Modeling Income in the Near Term - projections of retirement income through 2020 for the 1931-1960 birth cohorts. The Urban Institute, Washington D.C. Cited on pages 123 and 125.

Torssander, J. and R. Erikson (2009): Marital partner and mortality: the effects of social position of both spouses. *Journal of Epidemiology and Community Health*, 63:992–998. Cited on pages 162 and 181.

Train, K. (2003): *Discrete choice methods with simulation*. Cambridge University Press. Cited on page 74.

Trimp, L. (2000): Inkomensverdeling 1959-1998. In B. Grubben, J. Pelsers, and G. Timmermans, editors, *Jaarboek welvaartsverdeling 2000*, chapter 5, pages 41–53. Centraal bureau voor de Statistiek, Voorburg/Heerlen. Cited on page 112.

Tsai, S., J. Weendt, R. Donnelly, G. de Jong, and F. Ahmed (2005): Age at retirement and long-term survival of an industrial population: a prospective cohort study. *British Medical Journal*, 331:995–999. Cited on page 40.

Van Agtmaal-Wobma, E. and C. van Duin (2007): Huishoudensprognose 2006-2050: belangrijkste uitkomsten. *Bevolkingstrends*, 55(2):53–59. Cited on page 30.

Van den Berg, G. (2001): Duration models: Specification, identification, and multiple durations. In J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 5. North-Holland, Amsterdam. Cited on pages 42, 49, 162, and 170.

——— (2005): Competing risks models. IFAU working paper, no. 2005-25, Institute for Labour Market Policy Evaluation. Cited on page 51.

Van den Berg, G., M. Lindeboom, and F. Portrait (2006): Economic conditions early in life and individual mortality. *The American Economic Review*, 96:290–302. Cited on pages 160 and 162.

Van der Meulen, A. and F. Janssen (2007): Achtergronden en berekeningswijzen van CBS-overlevingstafels. Bevolkingstrends, 3e kwartaal. Cited on page 173.

Van Ewijk, C., D. Draper, H. ter Rele, and E. Westerhout (2006): *Ageing and the sustainability of Dutch public finances*. CPB Netherlands Bureau for Economic Policy Analysis. Cited on pages 9, 34, 39, and 109.

Van Houtven, C. and E. Norton (2004): Informal care and health care use of older adults. *Journal of Health Economics*, 23(6):1159–1180. Cited on page 66.

Van Kippersluis, H., O. O'Donell, and E. Van Doorslaer (2011): Long run returns to education: does schooling lead to an extended old age? *Journal of Human Resources*. Forthcoming. Cited on page 161.

Van Kippersluis, H., O. O'Donnell, E. van Doorslaer, and T. van Ourti (2010): Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science & Medicine*, 70:428–438. Cited on page 49.

Van Oorschot, W. (2007): Narrowing pathways to early retirement in the Netherlands. *Journal of Poverty and Justice*, 15(247–255). Cited on page 39.

Van Soest, A. (1995): Structural models of family labor supply; a discrete choice approach. *The Journal of Human Resources*, 30:63–88. Cited on pages 69, 75, and 83.

Van Soest, A. and E. Stancanelli (2010): Does income taxation affect partners' household chores. Document de travail de OFCE. Cited on pages 70 and 83.

Van Sonsbeek, J. (2009): Micro simulations on the effects of ageing-related policy measures: The Social Affairs Department of the Netherlands Ageing and Pensions Model. *International Journal of Microsimulation, forthcoming*. Cited on page 110.

Vaupel, J. and A. Yashin (1999): Cancer rates over age, time, and place: insights from stochastic models of heterogeneous populations. Max Planck Institute for Demographic Research working paper, no. 1999-06. Cited on page 51.

Vendrik, M. (2001): Dynamics of a household norm in female labour supply. *Journal of Economic Dynamics and Control*, 27:823–841. Cited on page 15.

Von Gaudecker, H. and R. Scholz (2007): Differential mortality by lifetime earnings in Germany. *Demographic Research*, 17(4):83–108. Cited on pages 128 and 159.

Waite, L. and S. Harrison (1992): Keeping in touch: How women in mid-life allocate social contacts among kith and kin. *Social Forces*, 70:637–655. Cited on pages 67 and 88.

Wallman, T., H. Wedel, S. Johansson, A. Rosengren, E. Eriksson, L. Welin, and K. Svärdsudd (2006): The prognosis for individuals on disability retirement: An 18-year mortality follow-up study of 6887 men and women sampled from the general population. *BMC Public Health*, 6:103. Cited on page 41.

Whitehouse, E. and A. Zaidi (2008): Socio-economic differences in mortality. OECD Social, Employment and Migration Working Papers, No. 71, OECD Publishing. Cited on page 160.

Wolf, D. (2001): The role of microsimulation in longitudinal data analysis. *Canadian Studies in Population*, 28:165–179. Cited on page 126.

Wolf, D. and B. Soldo (1994): Married women's allocation of time to employment and care of elderly parents. *The Journal of Human Resources*, 29(4):1259–1276. Cited on page 68.

Wooldridge, J. (2001): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge Massachusetts. Cited on page 172.

Yashin, A., K. Manton, and E. Stallard (1986): Dependent competing risks: a stochastic process model. *Journal of Mathematical Biology*, 24:119–140. Cited on page 51.

Zaidi, A. and K. Rake (2001): Dynamic microsimulation models: a review and some lessons for SAGE. SAGE discussion paper 2, Simulating social policy in an ageing society. Cited on page 110.