FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Master's Thesis

Anne-Line Koch Helsø

Modelling Retirement with Heterogeneity

A semi-parametric estimation with push- and pull effects

Thesis for the Master degree in Mathematics-Economics

Department of Mathematical Sciences, University of Copenhagen

Academic advisors: Peter Stephensen and Thomas Høgholm Jørgensen

Submitted: May 4th 2015

1 Abstract

A retirement model defined within a push/pull framework allowing for heterogeneity in leisure preferences is estimated. The proposed estimation method applies non-parametric estimation techniques and the estimation results show evidence of bi-modular population distributions of the leisure preference parameter k.

The model is estimated separately for ten gender- and education specific groups. The results suggest that men to a larger extent are pushed, while women are pulled into retirement. The estimated distributions of the pull parameter k tend to be split in a large group with relatively low values of k and a smaller group with significantly higher k values.

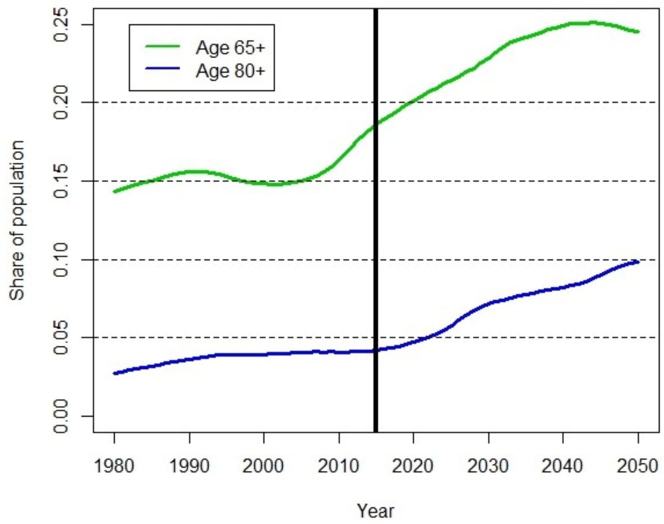
The estimated model assumes different interest rates such that the interest rate on deposits is constant while the interest rate on debt increases with the debt takers age. The Endogenous Grid Method (EGM) is applied to solve for optimal consumption paths. By explicitly accounting for data censoring, a potential selection bias and loss of useful information is avoided.

Contents

2 Introduction

Population aging is a pronounced and global trend, caused by rising life expectancies and falling birth rates. Figure 1 documents the persistence and magnitude of the trend in Denmark. The largest cohorts exit the work force while the smallest cohorts enters, resulting in increased dependency ratios and increased pressure on public finances. In 2014, Public spending on pension schemes accounted for 19.8% of Denmark's total public spending (Finansministeriet (2014)).

Figure 1: Historical and estimated development of elders in Denmark



Source: historical numbers and population projections from Statistics Denmark.

In addition to demographic changes, the share of Danes holding private pension wealth has increased significantly, making individuals more free to time their own retirement. About 4% of average wages were paid into to a pension scheme in 1984 compared to 11% in 2010. Conversely, payouts from individual pension savings rose from 15% of total pension payouts in 1984 to 35% in 2010 (Kramp et al. (2012)). An understanding of what constitutes people's retirement decision is crucial and the primary motivator for this analysis.

Structural retirement models are useful as they shed lights on the retirement decision and enables policy experiments and forecasts which are helpful tools for policy makers. The basic assumption when modeling behavior is that people's decision depends on a combination of their circumstances and their preferences. Most structural estimations assume that individuals share the same preferences, such that the variety in behavior is explained by the observed circumstances, e.g. financial situation, health etc. But given this approach, if a group of individuals facing the same circumstances behave differently, they must have different preferences. The purpose of this project is to measure and investigate this variety in individuals' preferences.

We model retirement behavior in a push/pull-framework that allows for heterogeneity in leisure preferences. The main contribution of this study is the proposed estimation of this heterogeneity which applies a non-parametric estimation technique. Another significant contribution is the actual estimation results that reveal a significant variation in peoples preferences for leisure, indicated by the model parameter k. The model is estimated separately for ten gender- and education specific groups. The estimated distributions of k tend to be split in a large group with relatively low values of k and a smaller group with significantly higher k values. This significant evidence of heterogeneity contributes with a deeper understanding of peoples retirement behavior.

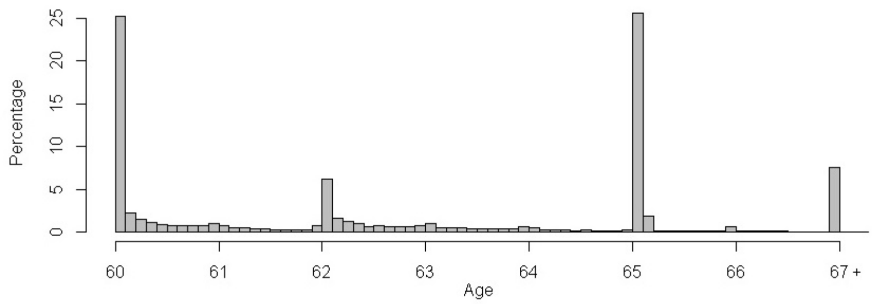
Chapter 3 outlines the context from which the current model has emerged and describes the applied model. **Chapter 4** describes the rules and regulations that constitute the Danish retirement system. The analysis is built on an extensive amount of data work as contra-factual income paths are simulated for all individuals, for each possible retirement age, described in **Chapter 5**. The model assumes different interest rates for debts (age-dependent) and deposits (constant), making a closed-form solution to the optimal consumption problem impossible. The consumption problem is therefore solved numerically with the Endogenous Gridpoint Method (EGM) proposed by Carroll (2006) in **Chapter 6**. The non-parametric estimation method is derived in **Chapter 7**. The estimation of the heterogeneous utility parameter is done together with the remaining homogeneous model parameters which are estimated with Maximum Likelihood. Finally, **Chapter 8** contains the estimation results that suggest bi-modularity in the population distribution of preferences of leisure and **Chapter 9** the final conclusions of the study.

3 Modeling Retirement

3.1 When and why do people retire?

In Denmark the "default" retirement age for those born prior to 1952 is 65 years - at age 65 you're eligible to receive old age pension (OAP). However, most people don't choose retire at age 65. Both early retirement schemes as well as private retirement savings enable people to retire prior to the default retirement age, while some choose to stay in the labor force after age 65. Figure 2 show a detailed histogram of the retirement ages of cohort $1941-45$ with peaks at age 60, 62 and 65. Retirement

Figure 2: Actual retirement ages, cohort 1941-45



Note: Source: Register data and the DREAM database (100% of the population). The week-precise retirement ages are provided in the DREAM database, retirement ages later than 65 are identified by the same criteria as in section 5.2.

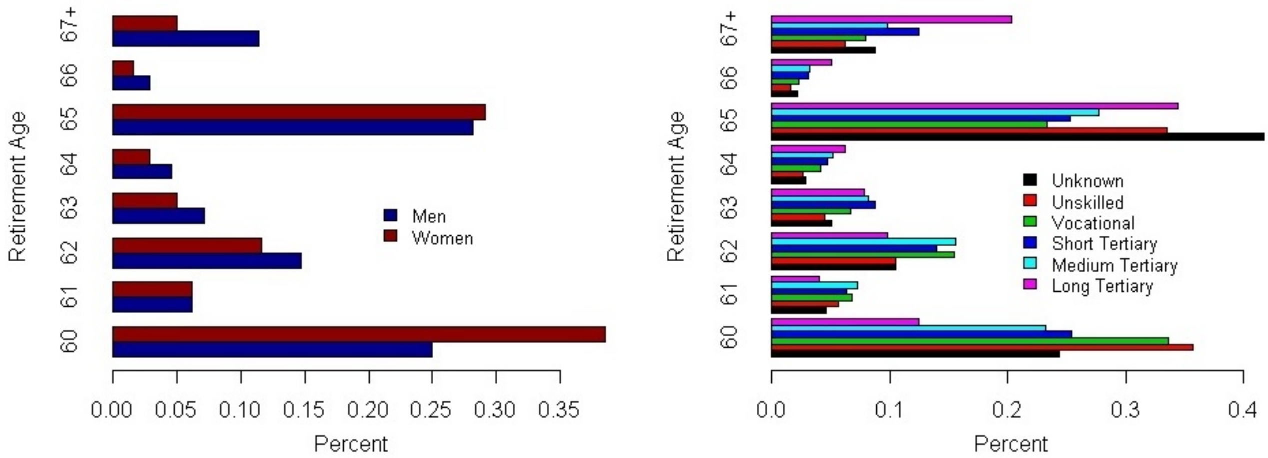
prior to age 65 must be funded either by own savings or, as is often the case, by Early Retirement Pension (ERP) benefits which enables retirement from age 60 for members of the scheme, see section 4.2. When we look at retirement ages for the different genders and educational groups (see figure 3), some patterns occur. Women tend to retire earlier than men and individuals with high education tend to retire later than those with less education.

The histograms reveal a standard with respect to retirement age in the Danish labor market. A large majority of Danes choose to retire between age 60 and 67. Obviously aging and attrition both plays significant roles in the matter, but what determines the *exact* retirement age of people during this age interval? The retirement decision of older workers depends on a great many variables. The following section outlines the most common reasons to retire, distinguishing between financial and non-financial determinants.

3.1.1 Financial Considerations

Economy is an important determinant in the retirement decision, consisting of savings, income and the underlying institutional framework such as the tax system and

Figure 3: Retirement ages for cohorts 1941-45 divided by gender and education



public pension schemes. The interaction between savings and income is ambiguous due to the nonlinear institutional settings. Financial circumstances improve when retirement is postponed. Because time as retired is shortened and extra time in the labor force enables individuals to accumulate more savings. In addition, public pension plans reward late retirement by increasing payouts and reducing offsets from private savings when retirement is postponed. Overall, the institutional framework encourages late retirement, but still leaves other options open.

Retirement often results in a significant decline in income. A low income while working results in a smaller decrease in income when retiring while high-earning individuals experience a larger decrease in income. Low wealth accumulations results in a low income while retired, encouraging late retirement. High accumulations offset the decline in income, encouraging early retirement. As a rule of thumb, combinations of a low labor income and high accumulation of savings make people retire early while high labor income and low accumulation of wealth make people retire late, see Table 1. It seems reasonable to assume that individuals with high salaries also have high

Table 1: How wealth and income affects the timing of retirement - Rule of thumb

	Low Wealth Accumulation	**High Wealth Accumulation**
Low Income	?	Early Retirement
High Income	Late Retirement	?

wealth accumulations and vice versa. Incentives related to the remaining two combinations are more complicated due to the complex and non-linear institutional settings. This complexity calls for a detailed assessment of each individual's financial situation in order to understand how his/her economic circumstances interacts with the timing of retirement.

3.1.2 Non-financial considerations

Were economy the only determinant, everyone would work until they droped dead. Obviously, there is more to the retirement decision than just economy, and the list of non-financial determinants could be infinitely long. Since the mid 90's, there has been a tradition within the field of retirement analysis to distinguish between pull and push effects of the retirement decision, e.g. Feldman (1994). While the push-effect covers negative features of the work role or working environment pushing people into retirement, the pull-effects cover the positive features of the retirement role or retirement environment pulling people into retirement. Push effects make people think that they *should* retire while the pull effects makes them *want* to retire.

- *Push:* The main reason why people are forced or pushed into retirement is their worsening **health** that often comes with aging, especially in jobs that require a high degree of physical activity. Having a physically demanding job might even contribute to the worsening health, why some experience physical **attrition** in their work, pushing them into retirement. Attrition could also be interpreted more broadly as the ability to keep up with the work, including the adjustment to new working procedures or new technology. Low job satisfaction, salary or job security as well as societal expectations of "normal" retirement age also adds to the list of common push-factors.
- *Pull:* Most individuals enjoy the **extra free time** that comes with retirement. A retired spouse, grandchildren or other factors that improve the value of free time will enlarge this effect. It is a standard assumption in economic theory that individuals value free time. But one could also imagine some people, who do not prefer free time over work. For older workers whose self-identity is closely tied to work, retirement means the loss of valued activities, not the gain of them.

Obviously, push and pull factors act concomitantly. It is the combination of pushes and pulls that produces preferences and intentions regarding the timing of retirement. Also, the extent to which the retirement decision is dominated by push forces as opposed to pull forces has consequences for the perception of retirement as a voluntary decision.

3.2 Literature study

Induced by a decreasing trend in participation rates of older workers, studies of economic determinants of retirement originated in the 1970ies in the US. In a context of low fertility and steadily increasing life expectancy, its progress has continued ever since. The following literature review will focus on the development of retirement studies in a global setting. A brief outline of Danish retirement studies will follow with an emphasis on more recent contributions. Even though all empirical analyses referred to below are based on different data sets and different methods, the results reveal some common features. All analyses suggest that financial circumstances affect the retirement decision. In addition, low education, spells of unemployment, and poor health are found to be associated with early retirement in all analyses where these issues are considered.

3.2.1 The evolution of retirement modeling

The earliest models of retirement are mainly applications of the standard **single-period labor supply model** where individuals maximize utility with respect to leisure and consumption. An example of this approach is Feldstein (1974). These rather primitive models assume that individuals solely focus on current period income when considering their retirement decision, ignoring the fact that future retirement benefits are affected as well. The one-period model was soon enough replaced with a **life-cycle model** where individuals act to optimize their discounted lifetime utility with respect to labor supply/retirement age. A lifetime retirement problem was already solved in Boskin (1977), and the continuous version of the problem can be generalized as:

$$\max_R U = \int_0^T U(L_t, C_t)e^{-\beta t}dt$$

$$s.t. \int_0^T C_t e^{-\beta t}dt = \int_0^R Y_t e^{-\beta t}dt + \int_R^T B_t(R)e^{-\beta t}dt$$

T denotes the maximum age of the individual, β the discount factor, Y the labor income and $B(R)$ the pension benefit retirement associated with retirement at time R. Empirical implementations of the life-cycle theory didn't take place until the 1980's. Mitchell and Fields (1984) and Gustman and Thomas (1986) provide two of the first empirical implementations of this general framework, modeling in detail the lifetime income pattern. The model in Gustman and Thomas (1986) was the first to explain the spikes of exits at age 62 and 65 in the US. Another frequently applied model of the 80's is the duration model, first developed for studies of unemployment spells, e.g. Hausman and Wise (1985) . The dependent variable of the duration model is time to retirement, taken as a positive, continuous random variable. But due to the reduced form nature of the duration model it lacks economic interpretation and predictability.

The introduction of **Dynamic Programming** (DP) was a significant contribution to the field of retirement studies when introduced in the late 1980's, first applied in Rust (1989) [1] At time t an individual can choose to 1) retire and derive utility from present and future pension or 2) to continue working and derive utility from current wage and the option to reconsider the problem in the following period:

$$V_t = \max\left[E_t\left(U_w(Y_t) + v_t + \beta V_{t+1}\right), \quad E_t\left(\sum_{s=t}^{T} \beta^{s-t}\left(U_r(B_s(t)) + \omega_s\right)\right)\right] \tag{1}$$

The utility function for workers U_w and a retirees U_R can vary and individuals are uncertain about future preferences which is reflected in the error terms v and ω which capture unobserved determinants of retirement such as health conditions. In order to define the retirement model, both functional forms of utilities and the structure of the error terms must be specified. The solution is obtained by recursive optimization (backwards induction). Dynamic Programming requires high computational complexity which increases when assumptions are relaxed. As technological improvement allowed for higher complexity, later DP models evolved towards more complex error structures. Stock and Wise (1990) assume a Markov Process for the error terms at the cost of "approximating" dynamic programming with the Option Value Model.

In the **Option Value Model** (OV) developed in Stock and Wise (1990), workers evaluate the following value function:

$$V_t(r) = \sum_{s=t}^{r-1} \beta^{s-t} U_w(Y_s) + \sum_{s=r}^{T} \beta^{s-t} U_R(B_s(r)) \tag{2}$$

The value function depends on retirement year r. In a two-stage comparison the worker determines what future retirement year yields the maximum expected utility (r^*), then he compares that utility ($E_t V_t(r^*)$) to the utility of retiring today ($E_t V_t(t)$). If the difference between these two (the so-called option value $G_t(r^*) = E_t V_t(r^*) - E_t V_t(t)$) is positive, the retirement decision is postponed to the following period, if $G_t(r^*) < 0$ he retires. Once retired, consumption is valued with parameter k compared to when retired, k representing the value of income while working relative to its value once retired.

The dynamic programming and option value models are much alike. They share the basic idea that workers decide whether to retire according to an evaluation of the opportunity cost in terms of utility. However they differ widely in the way they each treat uncertainty. The key simplifying assumption in the option value model is that the retirement decision is based on *the maximum of the expected* present values of future utilities if retirement occurs now versus at each of the potential future ages. The

[1] Rust (1987) first applied dynamic programming to solve for optimal replacement of bus engines. Rust's famous bus-example was one of the first and papers in the dynamic programming literature

DP rule considers instead the *expected value of the maximum* of current versus future options. The expected value of the maximum of a series of random variables will necessarily be greater than or equal to the maximum of the expected values. Thus, to the extent that this difference is significant, the option value rule understates the expected value of waiting. But the OV model has computational advantages over DP. Thanks to the absorbing state and Markovian assumptions, equation 2 can be divided into a deterministic and stochastic part, which drastically simplifies its computational complexity compared to dynamic programming, see Stock and Wise (1990). Moreover, Robin Lubsdain and Wise (1992) argues that *"more complex specifications may presume computational facility that is beyond the grasp of most real people and therefore less consistent with the actual rules that govern their behavior"*. And when comparing predictive abilities of dynamic programming with those of the option value, they find no significant difference. A more recent comparison in Belloni (2008) does not prefers one model to the other unambiguously either.

Econometric and technological advantages, together with availability of longitudinal data sets, allowed for increasingly complex models. The most prevalent models in recent retirement literature apply dynamic programming with modern specifications that relax many simplifying assumptions and includes the modeling of simultaneous decisions.

In the late 90s in the US a growing proportion of married women in the labor force approached old age. This led to a number of interesting questions regarding the phenomenon of **joint retirement**, i.e. spouses retiring simultaneously regardless of individual ages. Many studies have documented the joint retirement of couples, e.g. Blau (1998). Most models use the dynamic programming approach, one recent example being the quite comprehensive analysis in Gustman and Thomas (2009) which models the retirement decision on the household level, integrating many features of previous retirement models into a single framework. The model allows each spouse to retire and unretire, transitioning among the states of full-time work, partial retirement and full retirement. Retirement and savings are jointly determined and there is heterogeneity in time preferences. While most studies use the dynamic programming approach, some apply developments of the duration model, e.g. An et al. (1999) (Bivariate Duration Model) and Honoré and Paula (2013) (Interdepentent durations for husbands and wives).

Another prevalent subject in the field of retirement modeling is the interaction between retirement and **health**. Much of the available empirical evidence in this area suggests that poor health causes workers to retire earlier, e.g. McGarry (2004). Most studies have found that poor health brings on retirement but there is much disagreement as to the precise effects of health on retirement. The literature, however, does not unanimously agree on the size of the health effects - mainly due to the various econometric issues that arise when health is proxied by survey-based self-reports or even by more objectively measured indicators (Larsen and Gupta (2010)).

The model in Stock and Wise (1990) relies on assumptions about a 100% credit constraint with no saving or borrowing - consumption in a given period always equals income in the same period. More recent models relax this assumption and make the modeled **credit markets** more realistic. One example is French (2004) where individuals are allowed to save but not to borrow.

3.2.2 Recent Retirement Literature in Denmark

The Danish contribution to the retirement literature is quite modest, especially when it comes to structural models. Bingley et al. (2004) use the option value model to model the impact of financial incentives in the Danish retirement system and conduct a policy experiment. But instead of estimating the model on Danish data they simply apply the structural parameters found by Stock and Wise (1990). The first attempt to actually estimate the option value model in Danish data was done in Danø et al. (2004), where a subset of singles were used. The estimation was done separately on single men and women in order to compare their retirement behavior. The estimated values of the k parameters did not differ substantially from those estimated in Stock and Wise (1990) - they find k_{men} = 1.39 and $k_{females}$ = 1.37. However, they allow for k to vary with health and they expand their utility expression to include individual characteristics in an additive manner. They find that worse health increases the k-parameter and thereby individuals' willingness to reduce income in order to retire.

Analyses of joint retirement has also been prevalent in the Danish literature. A reduced form estimation of hazard rates can be found in An et al. (2004). Bingley and Lanot (2007) and Jørgensen (2014) apply a dynamic programming approach in which the individual retirement decision is modeled on the household level. Jørgensen (2014) finds that wives value leisure more than their husbands and low skilled workers value leisure more than high skilled. All three articles finds evidence of leisure complementarity in retirement for couples.

When it comes to health, Denmark differs from most other countries as health insurance is universal and access to most health services is free for all regardless of their economic situation. In addition, health-related exit from the labor market is possible in Denmark through Social Disability Pension (SDP, "førtidspension"). This implies that health ought to be less important in Denmark compared to studies conducted on e.g. US data. This theory is confirmed in Larsen and Gupta (2010). However, they too find health to be an important determinant of preferences for retirement in Denmark, with poor health status leading to earlier retirement.

Arnberg and Stephensen (2015) model retirement in a simple push/pull framework but allow for heterogeneity in preferences for leisure. This model constitutes the foundation of the current study and is described in detail in the next section.

3.3 The Model

Like the majority of its predecessors, the applied model defines how the retirement decision is based on preferences for consumption and leisure. This study is a development of the one in Arnberg and Stephensen (2015) which was originally developed in collaboration between the Danish Economic Councils and DREAM (Chapter 3 in "Dansk Økonomi Forår 2013"). The model is inspired by the option value model in Stock and Wise (1990), but differs significantly in the following points:

1. The parameter k which measures the utility of consumption obtained while retired relative to while working is assumed to vary in the population why we estimate a distribution of k's contrary to a single value.

2. Individuals know their future income streams and utilities with certainty.

3. Individuals can save and borrow - the interest rate on deposits is constant while the interest rate on debt increases with age.

4. The incentive to retire can increase with age as we include a variable that decreases the level of utility permanently as retirement is postponed. This parameter can be interpreted as an attrition parameter.

5. We apply a more general CRRA utility function with a higher coefficient of relative risk aversion, as is found in the literature.

One can say that simplifying assumptions about the decision process are made in order to allow for more complex specification of consumption and utility. Stock and Wise (1990) model uncertainty with respect to future utility functions in order to account for individual preferences for work versus leisure or evolving health status. The perfect foresight assumption in point 2 is better fit for this model as we explicitly account for individual specific leisure preferences and as the Danish welfare system is universal and free health care is available to all, why health shocks are less important in a Danish context compared to e.g. in the US.

The current analysis differs from the one in Arnberg and Stephensen (2015) by developing a more realistic capital market with different interest rates for deposits and savings. This makes an analytical closed-form solution to the consumer's problem impossible, why we instead solve it numerically by applying the Endogenous Grid Method (EGM) proposed by Carroll (2006). We also account for censored data which reduces a potential selection bias and increases the data size compared to Arnberg and Stephensen (2015). There are also some minor differences in the computing of future income streams, see section 5.3.

The model is derived in a random utility model framework in which all individuals are assumed to be utility maximizers. Consider N individuals at age 0 who are deciding when to retire, r. The time-subscript is denoted by a (age), $a \in \{0,1,...,A\}$ where A is the maximum possible age such that no individual lives longer than A years. Note that the time subscript is initiated in the year where the retirement decision is taken and not at birth, such that $a = 0$ might indicate an an individual at age 59. We assume that the individuals know their future financial situation - salaries, interest rates, pensions - with no uncertainty. The only uncertainty is with respect to death where age- and gender specific death probabilities are known to the individuals. The retirement age, r, is chosen at age 0 and retirement is considered an absorbing state - the retirement decision is assumed final and cannot be changed subsequently. Individuals choose to retire at age $1,...,P$ why all individuals aged P or older will be retired for sure. The utility at age 0 of retiring at age r is divided into a deterministic and stochastic component:

$$U_0(r) = \phi V_0(r) + \epsilon_r \tag{3}$$

V_0 denotes the present value utility of consumption in all future periods which is known to both the consumer and the researcher while ϵ_r is an unobserved heterogeneity known only by the consumer. Notice that ϵ applies for a given retirement age - it is not *age* dependent bur *retirement age* dependent. The non-stochastic term is called the representative utility. It is important to emphasize that ϵ_r is known to the consumer - from his point of view there is no uncertainty and the retirement decision is 100% deterministic. ϕ can be interpreted as a weight to the deterministic contribution relative to the utility function. As the variance of the stochastic contribution ϵ_r is exogenous, ϕ is identified. ϕ ensures that the size of the deterministic utility relative to ϵ_r makes sense. We could also interpret $\frac{1}{\phi}$ as the variance of ϵ_r - if ϕ is large the stochastic contribution ϵ_r will be small and the decision more deterministic.

Let u denote the temporary utility function in period a of period a consumption, c_a. All consumers are assumed to have CRRA utilities with relative risk parameter ρ which we will set to 2. But in addition to the traditional CRRA utility function, consumption is multiplied with the function $\gamma_a(r;k,\alpha)$ that introduces pull- and push parameters k and α related to retirement.

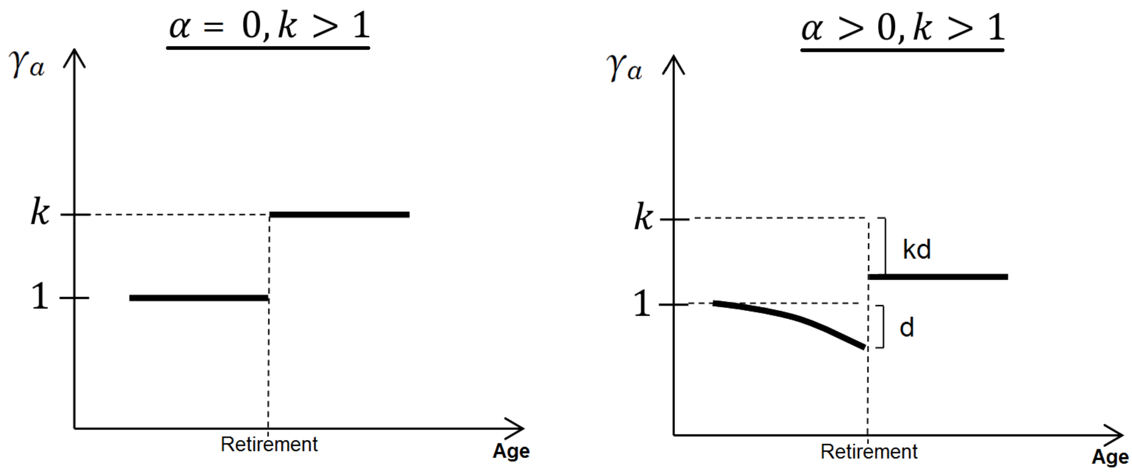
$$u(c_a) = \frac{(\gamma_a(r;k,\alpha)\, c_a)^{1-\rho}}{1-\rho} \tag{4}$$

γ_a is defined as

$$\gamma_a(r;k,\alpha) = \begin{cases} e^{-\alpha a^2} & \text{if } a < r \\ k e^{-\alpha r^2} & \text{if } a \geq r \end{cases} \tag{5}$$

The function $\gamma_a(r;k,\alpha)$ is drawn in figure 4 for both $\alpha = 0$ and $\alpha > 0$.

Figure 4: The function γ_a



While k is designed to capture the *pull* effects of retirement as described in section 3.1.2, α captures the effects that *pushes* people out of the labor force and into retirement. We assume that k is heterogeneous and we wish to estimate its distribution non-parametrically in the population. The effect of α is to decrease the utility gain of consumption prior to retirement. For every year retirement is postponed, the utility level decreases. Once retired, the utility level remains constant at the decreased value. k measures the utility effect of retirement and indicates how much a person values one unit of consumption after retirement relative to before. Setting $\alpha = 0$ we might also interpret $1/k$ as the fraction by which an individual is willing to reduce income in order to retire.

The present (age a=0) value of future discounted consumption, V_0, is given by:

$$V_0(r) = \sum_{a=1}^{A} \frac{\left(\gamma_a(r;k,\alpha)\, c_a\right)^{1-\rho}}{1-\rho} \beta_a \tag{6}$$

The discount factor β_a is defined as

$$\beta_a \equiv \prod_{i=1}^{a} \frac{1-\mu_i}{1+\theta} \tag{7}$$

where μ_a denotes the death probability at age a and θ the subjective rate of time preference. As the maximum age is A we have $\mu_A = 1$. An individual aged A years will

die within the year with probability 1.
The researcher observes the initial wealth W_0 and future incomes, $y_a, a = 0,...,A$ are observed or simulated from observed data. For $a = 0,...,r-1$ the income will consist of wages while for $a = r,...,A$ it will consist of retirement benefits. Let $Y(r)$ denote the income sequence given retirement age r, such that

$$Y(r) = \left(Y_1,...,Y_{r-1},B_r,B_{r+1},...,B_A\right)$$

with Y denoting income prior to - and B the income after - retirement. Given this information as well as assumptions about the utility function, interest rates, death probabilities and time preference, we are able to find an optimal consumption path as described in chapter 6. We assume that individuals choose to retire at the age, r, that gives the highest discounted utility at age 0, $U_0(r)$ of all future consumption. We do not know ϵ_r why it is considered a random variable.The error terms are assumed to be independent and identically extreme value distributed. As only choice-specific variables enters our utility measure the resulting model is a Conditional Logit model. More considerations on the choice of model as well as a mathematical derivation of the Logit model can be found in appendix 10.1. We can specify the probability of retirement at age r by

$$Pr\left(r|k,\alpha,\phi,\theta\right) = \frac{exp\left(\phi V_0(r)\right)}{\sum_{a=0}^{P} exp\left(\phi V_0(a)\right)} \tag{8}$$

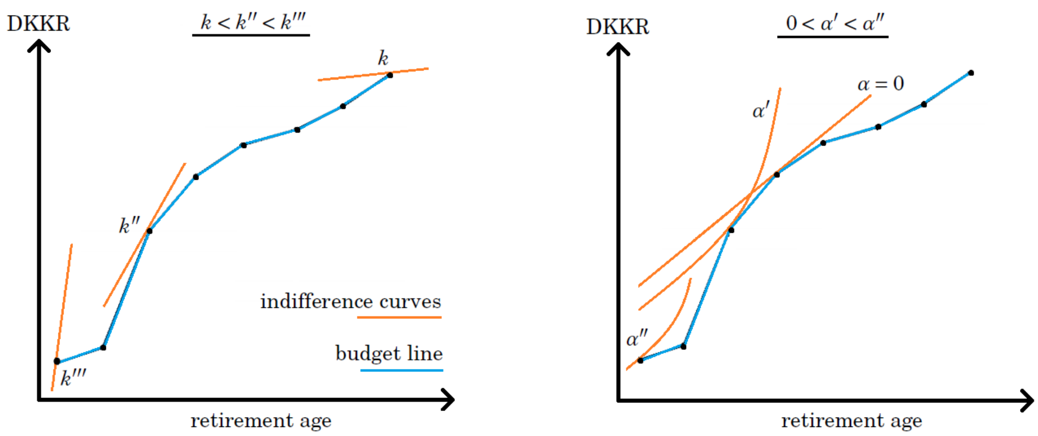
3.3.1 Model Intuition

Now consider the retirement decision from a classical consumer choice perspective with a budget line and indifference curves as illustrated in figure 5. The budget line indicates the present value of future incomes given the different retirement ages. Individuals typically wants high income and an early retirement age why they strive at reaching the top-left corner of the graph.

k affects the slope of the indifference curves and α the curvature. Individuals with high k values (high preferences for leisure) tend to cut the budget line in early retirement ages while those with low k-values cuts the budget line for late retirement ages. A high value of α increases the curvature of the indifference curve and make early retirement ages more attractive compared to when $\alpha = 0$. Some retirement ages are optimal for several values of k and α.

ϵ_r denotes the unobserved heterogeneity in the model and it covers non-financial incentives that affect the retirement decision. Interpretation of the error term is not straight-forward and its effect is easily confused with that of k. It is essential to recall that ϵ_r is retirement age dependent: An individual has P unobserved error terms - one

Figure 5: Optimal Retirement Age with (right) and without (left) push effect α



for each possible retirement age - which we assume to be identical and independently extreme value distributed, see Section 7.5 for further reflections on this assumption. Therefore we might think of ϵ_r as an error term that captures outer stochastic shocks relating to a given retirement age. On the other hand, k captures some deeper personality characteristic that is constant for an individual.

Recall that ϕ indicates how deterministic/stochastic the model is with $1/\phi$ scaling the variance of the error term ϵ_r. This is emphasized when evaluating equation 8 in the limit values of ϕ. If $\phi = 0$ vi get:

$$Pr\left(r|k,\alpha,\phi=0,\theta\right) = \frac{1}{P} \quad \forall r \in \{1,...,P\} \tag{9}$$

Such that the retirement decision is 100% random with equal probability of each retirement year. The opposite case is when $\phi \to \infty$:

$$\lim_{\phi\to\infty} Pr\left(r^*|k,\alpha,\phi,\theta\right) = 1 \tag{10}$$

With r^* denoting the utility-maximizing retirement age such that

$$r^* = \underset{r\leq P}{\text{argmax}} \quad V_0(r)$$

If ϕ is infinitely high, r^* would be the chosen retirement age with certainty. A small ϕ indicates that the unobserved heterogeneity has a great impact on the retirement decision, ϵ_r being unknown to the researcher and assumed i.i.d extreme value distributed.

As k increases, individuals are more likely to retire early, and when k approached 0 they are more likely to retire late. However, due to the error term ϵ_r and the fact that $\rho = 2$, an infinitely large k does not implicate retirement at $r = 1$ with certainty. For a further elaboration on the properties of equation 8 with respect to k, see Section 7.1.2.

We set $\rho = 2$. Recall that ρ denotes the coefficient of relative risk aversion in the assumed CRRA utility function, $1/\rho$ being the inter-temporal elasticity of substitution (IES). $\rho = 0$ corresponds to an additive utility function where the consumer is indifferent between consumption now and in the future. $\rho = \infty$ corresponds to a limitational utility function where the consumer doesn't accept any variation in consumption and choose a constant consumption path. General speaking, $\rho < 1$ indicates low preferences of consumption smoothing while $\rho > 1$ indicates high preferences of consumption smoothing. The value of ρ also has an effect on the curvature of the budget line in Figure 5 and obviously influences the estimation results. However, Arnberg and Stephensen (2013) argues that the effect is rather small. Previous literature on risk aversion does not agree on the estimate of ρ, which varies from 1 to 10 in most studies (Azar (2010)). A general consensus is to put $\rho = 2$, as this is done in Arnberg and Stephensen (2015). We do the same in order to keep our estimates comparable.

Great caution is needed when interpreting the different parameter values. The push- and pull variables α and k are modeled such that α captures a worsening of the utility level as retirement is postponed, a decline that continues until retirement and then stabilizes at a permanently lowered level. k on the other hand, reflects a jump in utility after retirement. It is constant and does not depend on the chosen retirement age. The push and pull effects described in section 3.1.2 does not all fit equally well into the chosen modeling. Some push effects, e.g. the ability to keep up with work, job insecurity or general dissatisfaction with work does not result in a permanently decrease in utility after retirement. Job dissatisfaction, e.g. is better captured in the push parameter as a dislike of working is also reflected in higher preferences for leisure.

As the model doesn't account explicitly for retirement-related variables such as health, it is difficult to relate the estimated push/pull variables directly to e.g. health conditions. If an individuals' bad health is worsened when staying in the labor force (e.g. physical attrition) it would show in a large pull parameter α. If an individuals' bad health condition on the other hand is permanent and not worsening with age (e.g. an individual in a wheel chair) it might show in higher preferences for leisure, k. A more temporary bad health condition predicTable at age $a = 0$, such as the knowledge of a future knee operation, might again show in ϵ_r.The aim of the specified model is not to identify the effect of certain circumstances as health or joint retirement. Its advantage, on the other hand, is its very simplistic and pure utility framework that cuts to the actual core of the retirement decision. The fact that the model only includes future income streams as explanatory variables makes it ideal for out-of-sample forecasts but also makes it lack on direct economic interpretation.

4 The Danish Retirement System

Institutional settings play a key role when seniors form their labor market decisions. The following section outlines the Danish retirement system that applies to the cohorts of 1941-45 with an emphasis on the rules that creates incentives with respect to the timing of retirement. The two main elements of the Danish retirement systems are the *Early Retirement Pension*, ERP (efterløn), and *Old Age Pension*, OAP (folkepension). ERP is a voluntary program in which participants pay annual membership fees in order to obtain eligibility from age 60 to 65. OAP is fully government financed and available to all Danish citizens at the age of 65. To supplement the ERP/OAP benefits, individuals can accumulate *individual pension savings*. Section 5.3 includes a more detailed description of how different institutional settings are implemented in the computation of income simulations.

4.1 Individual pension savings

For the majority of Danes, living off OAP alone will result in a drastic decline in income after retirement, why increasingly many accumulate individual retirement savings as a supplement to the OAP. The private pensions in Denmark are divided into *two main categories*: employer and employee administrated savings (in Danish arbejdsgiveradministrerede and private opsparinger). Employee administrated pensions are agreement based schemes where employers contribute a percentage of the worker's gross labor market income to an individual pension fund, typically ranging from 12 to 17 percent. These are typically decided through collective agreements. Employee-administrated pension funds consists of savings that are accumulated voluntary by the employee.

The two main categories (employer and employee administrated savings) can be further divided into 3 **types** of retirement savings: life annuities ("livrente"), term pension ("ratepension") and capital pension ("kapitalpension"). First, *life annuities* guarantees a monthly payment from retirement until death. The guaranteed amount depends on the level of accumulated savings and is therefore increased (decreased) if the owner postpone (advance) retirement. Secondly, *term pension* is a pension balance that is distributed through annuities of 10 through 25 years, initiated no later than age 77. Thirdly, *capital pension* is a pension balance with no requirements on installment - is usually paid as a lump sump, no longer than 15 years after retirement.

4.2 Early Retirement Pension (ERP)

The ERP system was introduced in the 1970'ies with the purpose to make room for new and younger employees by allowing the old and physically attired workers to retire. ERP benefits are available from age 60 until the default retirement age at 65, where the OAP is available and contains elements of both a funded and unfunded benefit system. Members of unemployment funds with a certain degree of seniority, who have contributed to the ERP program for at least 10 years prior to retirement (yearly fee was 4.488 DKK in 2004) are entitled to receive ERP benefits. The ERP payout resembles the level of unemployment benefits, making the ERP scheme very popular as the government roughly finances 70% of the benefits (Jørgensen (2014)). To qualify for ERP, one must be available to the labor market, such that individuals retiring prior to age 60 loose their ERP eligibility. This requirement of labor market attachment makes retirement prior to age 60 very unattractive and results in a significant spike of retirements at age 60. Approximately 72% of the cohorts 1941-45 were ERP eligible at age 60 [2], making the eligible share of those attached to the labor market even higher (92% in this analysis, see chapter 5). A so-called *two-year rule* makes it more attractive to postpone ERP retirement for two years, until age 62. If an individual has been eligible for ERP in at least two years before retiring, the level of ERP increases from 91% of the unemployment benefits to 100% [3]. The ERP benefits are means tested with respect to individual pension savings, depending on the level, type (life annuity, rate or capital pension) and category (employer or employee administrated) of the savings, and wheter if the pension is paid concurrently with the ERP benefits. Fulfilling the two-year rule will mean a decrease in ERP benefit deductions. How the ERP is means tested with respect to the different types of individual pension savings is described in detail in Table 2. When an ERP eligible individual fulfills the two-year-rule but postpones his retirement even further, he receives one "portion" tax-free premium of 10,000 DKK (2004-level) for every 4 months of full-time work until age 65, making the maximum possible number of portions equal 12.

4.3 Old Age Pension (OAP)

The Danish OAP system is available to all Danish citizens aged 65 or above. It consists of a baseline amount ("grundbeløb") and an OAP supplement ("pensionstillæg").The baseline annual amount (55.776 DKK in 2004) is means tested based on concurrent labor market income only, while the additional amount differs for married (26.208 DKK

[2]Source: Register data. An individual is assumed ERP eligible if he/she 1) Contributed to the ERP program at age 59 or 2) Received ERP benefits

[3]The level of unemployment benefits is set to 90% of the previous salary but with a maximum limit (166.660 DKK in 2004)

Table 2: Annual Means Testing of ERP based on individual pension savings

		Retirement Before 62		Retirement After 62	
		Employer Adm.	*Employee Adm.*	*Employer Adm.*	*Employee Adm.*
Life Annuities	*NP*	60% of (80% of RP - BA)	60% of (80% of RP - BA)	0	0
	P	50% of AP	60% of (80% of RP - BA)	55% of AP	0
Term Pension	*NP*	60% of (5% of RD - BA)	60% of (5% of RD - BA)	0	0
	P	50% of AP	60% of (80% of RP - BA)	55% of AP	0
Capital Pension		60% of (5% of RD - BA)	60% of (5% of RD - BA)	0	0

Note: NP = No Payments Made, **P** = Payments Made, **AP** = Actual Payment (annual), **RP** = Reported Payment (annual), **RD** = Reported Deposited amount of total savings, **BA** = Basic Allowance (11.500 DKK in 2004) - can only be used once.

in 2004) and unmarried (56.148 DKK in 2004) and depends on a number of factors including individual pension savings, concurrent earnings of a partner and his/her labor market status. Unlike the ERP, OAP is *not* means tested based on pension savings. Seniors with poor financial situations can furthermore apply for additional benefits, e.g. housing allowances. If an individual at age 65 postpones his retirement, his OAP will increase with 6% for each year of postponement (with a maximum of 10 years). If e.g. the OAP is postponed with 2, 3 or 4 years, OAP benefits will increase permanently with 12, 18 and 24%.

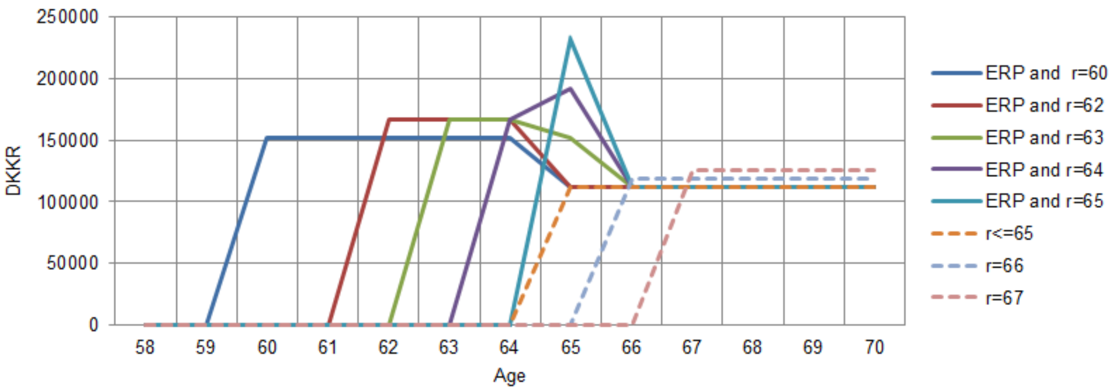
4.4 Acting on financial incentives

While favoring late retirement, the Danish retirement system induces retirement at ages 60 (ERP), 62 (ERP/private pensions) and 65 (OAP). One essential assumption underlying the entire study is that the retirement decision depends on financial incentives - if not, the analysis would be completely nonsensical. Fortunately, people do act on financial incentives when timing their retirement. Danish register data for cohorts 1941-45 show that most people retires when there is a financial incentive. Figure 2 in section 3.1 depicts the actual retirement ages for cohorts 1941-45, and we see that actual retirement ages cluster at age 60, 62 and 65.

Figure 6 reports the age-dependent paths of ERP and OAP benefits for different combinations of ERP eligibility and retirement ages. The 2004-rates applies to a single person with no individual pension savings and no labor market income in retirement.

The solid curves apply to an ERP eligible person who retires at ages 60 (low level of ERP), 62 (high level of ERP), 63/64 (high level of ERP + 4/8 taxfree premium portions) and 65 (no ERP benefits but 12 tax free premium portions). The dotted lines apply to a person who is not eligible of ERP where retirement prior to age 65 must be self-financed. Recall that an individual eligible to ERP who retires at age 59 looses his ERP elegibility. Retirement later than age 65 result in a permanent 6% increase in OAP benefits for each year the retirement is postponed. This, off course, also applies to ERP eligible individuals who would then be rewarded with a tax free premium *and* increased OAP.

Figure 6: Annual retirement benefits from ERP and OAP



Note: Figure 6 illustrates how the ERP and OAP scheme depends on age for different combinations of ERP eligibility and retirement age r. The illustration is based on a single individual with no private pension wealth. To simplify, we have only used 2004-rates and disregarded any discounting.

5 Data

The analysis is based on a vast amount of population-based administrative register data available from Statistics Denmark. The data contains anonymised information on a 33.3% random sample of all Danish citizens. Individuals born between 1941 and 1945 are selected form age 59 to 67. Danish register data includes yearly information about earnings, savings, social transfers etc. The initial data processing is performed with SAS software while taxes and public transfers is modeled in *C#*.

5.1 Scope of the model

The model is estimated with first possible retirement age $1 = 60$ and last $P = 67$, such that the decision age is $0 = 59$. The majority of the cohorts 1941 to 1945 retires during this interval of time, see figure 2. However, approximately 7% retire at age 67 or later. It would always be optimal to prolong the period of possible retirement ages, but this would increase the amount of data work and estimation computations considerably. As a compromise, 67 seems to be a reasonable choice of latest retirement age. The same argument goes with the choice of maximum age which is set to 100 years.

5.2 Data selection

The selection of data has a huge impact on the following empirical analysis why we must be careful to select our data in a reasonable manner. As the response variable is retired/not retired it only makes sense to include those individuals who face a "standard" retirement decision, why individuals on disability- and transition benefits together with officials [4] are excluded. Table 3 outlines how the data size decreases when observations are discarded, ending up with roughly 40% of the original data. 91,7% of our sample is entitled to the ERP.

Table 3: Data Selection

	Total	Deleted	Share
Cohorts 1941-1945 alive at age 59	107932		
Salary >90,000 (2001-level) at age 59	70405	37527	34.77%
Haven't been on disability or transition benefits	68927	1478	1.37%
Isn't an official	65581	3346	3.10%
Total	**65581**	**42351**	**39.24%**

[4]Officials refer to those eligible for the Danish equivalence of a defined benefit plan in the US ("tjenestemandspension")

Individuals who are not present in data throughout all 8 consecutive years (age 60 to 67) are not discarded from the analysis. Section 7.3 how the data censoring are handled throughout the estimation process. Including the "disappearing" individuals in the analysis minimizes a potential selection bias. Note that self-employed are included in the analysis. One might argue that the retirement decision of self employed depends on other factors than for employees, but assuming an overall similar push/pull structure seems quite reasonable.

Our data is divided into 12 gender/education sub-samples prior to estimation. Table 4 show the sizes of these groups and how income, wealth and pension savings vary through the different groups. It also show how the majority of the sample are eligible for ERP, corresponding to roughly 90% of the entire sample, with a small over representation of women. We see that males have higher labor market incomes and considerably higher wealth accumulations (more than double in mean and quadrupole in median) compared to their female counterparts. The male observations also have larger pension savings than females, but much more equally distributed between the genders than wealth accumulations. Another prevalent tendency for both earnings, wealth accumulations and pension savings is that they increase with education.

Table 4: Descriptive Statistics

	N	ERP %	$\overline{Y}_{59}$	$\tilde{Y}_{59}$	$\overline{W}_{59}$	$\tilde{W}_{59}$	$\overline{PS}_{59}$	$\tilde{PS}_{59}$
MEN	**36,130**	**87.83**	**366**	**302**	**853**	**504**	**1056**	**288**
Unknown	570	80.18	346	286	673	331	889	253
Unskilled	10,709	88.73	327	271	812	422	666	247
Vocational	15,787	91.49	331	288	741	475	646	261
Short Tertiary	1,538	83.62	365	322	1051	651	746	256
Medium Tertiary	4,884	84.50	439	372	962	647	1,886	589
Long Tertiary	2,642	72.63	595	475	1,413	877	3,759	3,682
WOMEN	**29,454**	**91.95**	**252**	**236**	**399**	**125**	**843**	**246**
Unknown	408	83.33	266	256	505	162	827	261
Unskilled	10,201	92.01	223	212	311	79	426	197
Vocational	11,417	92.87	241	231	353	120	581	237
Short Tertiary	1,012	89.13	276	260	613	251	921	344
Medium Tertiary	5,347	93.27	293	293	499	194	1,741	904
Long Tertiary	1,069	81.01	430	401	974	624	3,002	2,443

Note: All amounts are in 1.000 DKK. Bar denotes mean value, while tilde denotes the median. Y_{59} is labor market income at 59 years, W_{59} is wealth at age 59 (initial wealth) while PS_{59} is the total amount of deposited pension savings, including life annuities, term pensions and capital pensions. ERP % denotes the percentage of a given group that are eligible for ERP benefits.

An individual is defined to be retired if at least one of the below statements are true: 1) Receives more than 68,000 DKK/year (2000-level) from individual pension savings, 2) Receives ERP benefits, 3) Receives OAP benefits with a yearly salary below 90,000 DKK (2000-level), 4) Receives OAP or ERP benefits in 1-11 months and receives less than 50% of previous year's salary or 5) Yearly salary below 90,000 DKK (2000-level) two years in a row. The distribution of retirement ages in the sample, including the gender-specific distributions, are stated in Table 5.

Table 5: Gender-specific distribution of educational attainment

	All	**Men**	**Women**
60	24.7%	18.9%	32.0%
61	9.6%	8.3%	11.4%
62	18.0%	17.9%	18.1%
63	9.9%	10.2%	9.4%
64	6.3%	7.0%	5.5%
65	10.9%	12.5%	8.9%
66	5.1%	5.7%	4.3%
67+	13.5%	16.9%	9.2%
Missing	2.0%	2.6%	1.2%

5.3 Simulation of future income streams

An extensive amount of data work is processed in order to compute the future income streams. For each individual, income at all ages $a \in \{60,...,100\}$ must be computed for all possible retirement ages, $r \in \{60,...,67\}$. The future income streams consist of salaries, public retirement benefits (ERP/OAP) and payments from individual pension savings. Some parts of the income streams are read directly in data, but far the majority of future incomes are simulated, as they relate to a hypothetical retirement ages or simply happen in the future. We are able to observe the individuals' retirement savings at age 59 together with salaries and wealth accumulations. The future income streams consists of salaries, ERP benefits, OAP benefits and payments from individual pension savings. The following section will explain the assumptions that are made in order to simulations of the different components. ERP, OAP and taxes depend on a highly complex system of rules. They are simulated in a comprehensive tax-benefit C# simulator developed by DREAM.

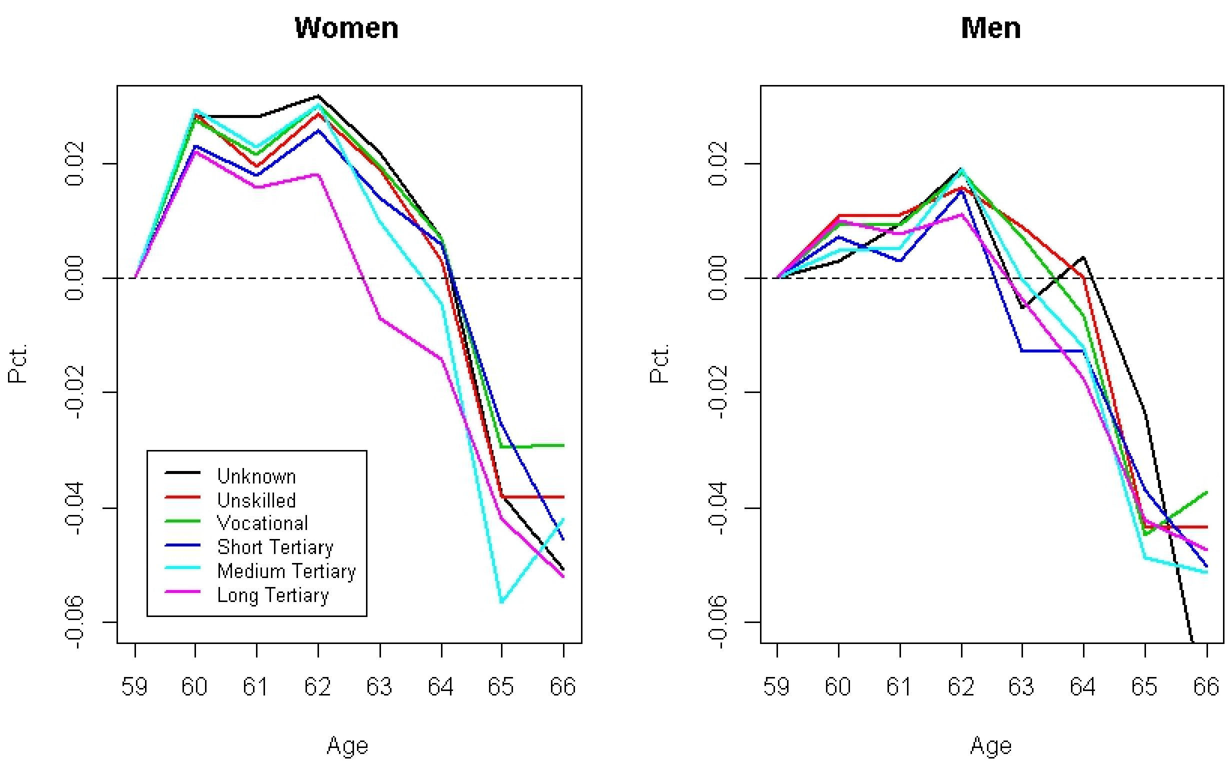
5.3.1 Wealth

Individual wealth is set to the variable *formrest_ny05* - constructed by Statistics Denmark - observed at age 59. It includes property value, bank deposits, shares, bonds and mortgages deducted debts in different financial institutions including mortgage and bank debt. While this variable covers the most important elements of wealth it doesn't include value of cars, boats, cash and share purchases in cooperative housing. We let the largest initial debts equal 1,000,000 DKK in order to facilitate optimal consumption paths with non negative consumption.

5.3.2 Salaries

Annual salaries are observed for both wage earners and self-employed for those years they actually work. Hypothetical salaries from individuals' actual retirement age until the last possible retirement age (67) must be simulated. The chosen simulation is indeed very simple: we compute the average age-specific wage development for each gender-education group from observed salaries and forecast the simulated salaries with the same development. The simulated wage of individual X subsequent to his actual retirement is found by assuming that his wage would have followed the same development as that in his corresponding gender/educational group, had he not retired. The two graphs in figure 7 depict the developments for the different gender- and educational groups. After age 60 wages increases during the first couple of years while decreasing around age 63-64 for all groups. The wage increase is higher for women than men.

Figure 7: Wage Development



5.3.3 Early Retirement Pensions (ERP)

We assume that all individuals contributing to the ERP-scheme at age 59 are entitled to receive ERP benefits at age 60 (roughly 90% of our sample, see Table 4) and that they meet the 2-year rule at age 62. We assume all individuals to be full-time ensured. Table 2 in section 4.2 outlines how the ERP benefit is means tested with respect to retirement savings. Individuals who retires after age 62, thereby fulfilling the two-year-rule, are assumed to postpone any individual pension payments until age 65 in order to avoid reductions in the ERP benefits. If a person fulfills the 2-year rule but postpones retirement, he will receive a tax-free bonus of approximately 40,000 DKK for each year of postponement.

5.3.4 Old Age Pension (OAP)

We assume that all individuals older than 65 are entitled to receive the OAP. The basic OAP will only be reduced if the individual receives a salary of a certain amount. The amount of maximum additional OAP depends on cohabiting partner status and the amount is reduced with the level of total income (including private pension payouts), partner's income and retirement status. The different rates for the OAP are set to the actual rates until 2015. Future rates are assumed to grow with an assumed inflation rate of 3.28%. Individuals who retires at age 66 will receive 6% increased OAP benefits for the rest of his life. Postponing retirement until age 67 will increase it by 12%. This incentive to postpone retirement was not introduced until 2004, but it is included for all individuals. This is in fact a violation of the model assumption, see further explanation in section 7.5.

5.3.5 Individual Retirement Savings

Information about all individuals' private retirement savings are reported at age 59.5 by regulation. While capital pensions and term pensions are both listed with the deposited amount, life annuities are often registered with the annual commitment giving retirement at age 60. Individuals are assumed to save some percentage of their wage for their retirement as long as they work. Information about actual yearly pension contributions are available from data until the actual retirement age. The mean shares of income that was paid to the different types of retirement savings from age 59 until the actual retirement age are computed. We assume that the same share of income would be paid in the years following the actual retirement age in order to simulate the fictive contributions.

5.3.5.1 Capital Pension

We assume that the Capital Pension (CP) is paid as a lump sum in the first year of retirement. The capital pension deposit are assumed to grow with the annual interest rate $i_d = 4.75\%$ in the period prior to retirement. All interest gains on retirement savings are taxed with the so-called PAL tax, $\tau_{PAL} = 15.3\%$. CP_{59} is observed from data and the subsequent years are found by:

$$CP_a = CP_{a-1} * (1 + i_d(1 - \tau_{PAL})) + \Delta CP_{a-1} \text{ for } a \leq r$$

where ΔCP_{a-1} denotes the amount contributed to the capital pension at age $a-1$

5.3.5.2 Term Pension

All terms pensions are assumed to be equally distributed through annuities of 10 years, such that the payment size simply equals 10% of the deposited amount at age the retirement age, growing with $1 + i_r(1 - \tau_{PAL})$ each year. The payments start at the year of retirement with exemption of ERP eligible individuals who retire at age 63 or 64. They are assumed to postpone the payments until age 65 to avoid reductions in ERP benefits.

5.3.5.3 Life annuities

Life annuities are observed both as total commitment (LA^{TOT}) and as annual payouts (LA^{PAY}) given retirement at age 60. In order accumulate the contributions made after age 59 (ΔLA^{TOT}), all life annuities must be transformed into their corresponding total deposited values. And then, in order to compute the life annuity payments in retirement, the total deposited amount must be transformed back into the corresponding annual payments. These transformations are done with some actuarial mathematics. Assuming that the deposited values at all times should equal the present value of future annuities we get the following:

$$LA_a^{PAY} = \frac{LA_a^{TOT}}{\sum_{i=a}^{100} \left(1 - \mu_i\right) \times \left(\frac{1+i_d(1-\tau_{PAL})}{1+\pi_{wage}}\right)^{-(i-a+1)}}$$

The total committed amount, LA^{TOT}, is assumed to follow same development as the capital pension such that

$$LA_a^{TOT} = LA_{a-1}^{TOT} * (1 + i_d(1 - \tau_{PAL})) + \Delta LA_{a-1}^{TOT} \text{ for } a \leq r$$

where ΔLA_{a-1}^{TOT} denotes the amount contributed to the life annuity savings at age $a-1$. LA_a^{TOT} denotes the age a fixed-price value of the total commitment and μ_i the death probability at age i. The interests gained on the deposited value are assumed to equal the interest rate on deposits $i_d = 4.75\%$. We let π_{wage} denote the wage inflation, set to growth (g=1.5%) times price inflation $\pi_{price} = 1.75\%$, such that $\pi_{wage} = 1.015 * 1.0175 = 1.032 = 3.2\%$.

5.3.6 Spouses

Now consider *the main person* and his/her spouse - *the spouse*. As the income and retirement status of the spouse affect the amount of OAP which the main character is entitled to, we need to compute spouse salaries. The exact retirement age of the spouse is not necessarily observed. In that case, it is set to the default retirement age, 65. The latest observed salary of the spouse is extrapolated with wage inflation $\pi_{wage} = 3.2\%$ until the actual or assumed retirement. Some individuals loose their spouse and/or get a new spouse during the age interval 60 to 67. You can argue both against and in favor of including observed changes in partner status when computing the future income streams. Whether individuals are able to predict divorces, death of a partner, meeting a new partner etc. is a delicate matter. We decide to include all observed changes in partner status with reference to section 7.5.3 on the perfect foresight assumption.

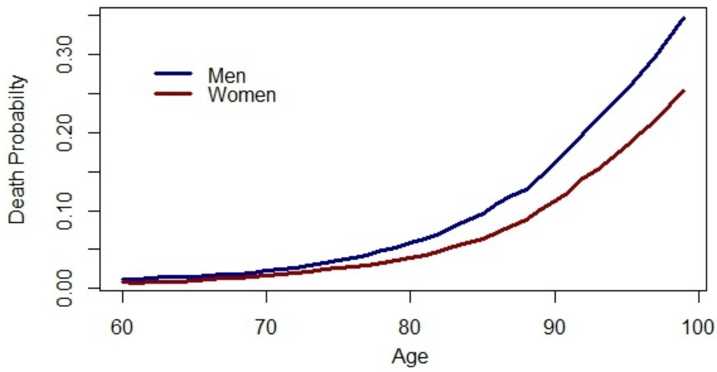
5.3.7 The Tax System

Individuals are, at age 59, obviously not able to predict how the tax system develop in the future. Never the less, actual tax institutions are applied until and including 2014. Thereafter we will "freeze" the overall tax system and only project the different limit amounts for the different progressive tax-levels with the wage inflation $\pi_{wage} = 3.2\%$. One might argue that it would be more correct to freeze the tax system from age 59, but this would result in drawbacks with respect to those individuals who retire late and include changes in the taxsystem before retirement, see section 7.5.3.

5.3.8 Mortalities

We apply the gender- and age specific death probabilities calculated by DREAM for the 1943 cohort. As we assume that the maximum age is 100 we have put $\mu_{100} = 1$. The age and gender-specific mortality for age 60-99 are depicted below in figure 8, representing the probability of dying in the following year at a given age. As is well known, mortality is highest for men and increases with age.

Figure 8: Death Probabilities



5.4 Omissions

Due to data- and time limitations some transfers and incomes were excluded in the analysis. All capital gains from the initial wealth accumulation is set to zero as an estimation of future capital gains would include estimation of future wealth which is found endogenously in the model. Besides contributing to the future income stream, capital gains also affect the amount of additional OAP and taxes. These effects are not captured in our computed income streams. Nor do we include housing benefits in our calculations. The amount of entitled housing benefits depends on the tenant's financial situation (both income and wealth) and the amount is higher for those eligible of OAP. It would be quite naive to assume that 59-year olds remain in the same residence until their death, why an additional model of future moving patterns should be included in order to find a reasonable suggestion of future housing benefits. Due to the high complexity of the housing benefits it is further more arguable that most individuals are not able to forecast their own future amount of housing benefits. Other public benefits targeting the economically disadvantaged are excluded as well. The amount depends on the individual's future wealth which, as just mentioned, is endogenously determined in the model. These benefits are the supplementary pension benefits (ældrecheck), health service supplements (helbredstillæg), heating allowance (varmetillæg) and discounted TV license.

6 Optimal Consumption Paths

An essential assumption when analyzing the retirement decision lies in the individuals' expectations of the future. How they act today depends on their beliefs about tomorrow. A broad agreement in previous literature on the subject is to consider the retirement decision from a life-cycle perspective. The consumer chooses his levels of consumption, savings and work supply such that the total expected utility of the future is maximized. In this analysis we assume that individuals have perfect foresight with respect to their future financial situation. As a result, the consumer is able to derive a final future optimal consumption path at $a = 0$.

6.1 Optimization problem with credit constrains

Our attempt to model a realistic credit market is described in the following: The consumer is able to save and lend money at all times, but is subject to a credit constraint consisting of varying interest rates on debt and deposits. The interest rate received from deposits (i_b, "b" for bonds) will always be smaller than the interests on debt (i_d). The relationship between the two interest rates is defined as:

$$1 + i_b = \left(1 + i_d(a)\right)\left(1 - \mu_a\right) \tag{11}$$

While the interest rate on deposits/bonds is constant, the interest on debt varies with the debt takers age: $1 + i_d(a) = \frac{(1+i_b)}{(1-\mu_a)}$. When deciding upon an interest rate, the moneylender takes the money borrower's death probability into account in order to deal with the risk of the borrower dying and never paying back the loan, making it an actuarial fair contract.

We denote the end-of-period net wealth by W_a, start-of-period income by y_a and the during-period consumption by c_a. W_a can attain both positive (deposits) and negative (debts) values. As described above, the interest rates accumulated from the previous end-of-period wealth depends on its sign:

$$\hat{i}(a) = \psi_a(W_{a-1}) = \begin{cases} i_d(a) & \text{if } W_{a-1} < 0 \\ i_b & \text{if } W_{a-1} \geq 0 \end{cases} \tag{12}$$

Given initial wealth (W_0) and future income streams ($y = \{y_1, ..., y_A\}$), the $a = 0$-year old faces the following maximization problem with respect to consumption path $c = \{c_1, ..., c_A\}$:

$$\max_c \sum_{a=0}^{T} \frac{\left(\gamma_a\left(r; k, \alpha\right) c_s\right)^{1-\rho}}{1-\rho} \beta_a \tag{13}$$

Subject to the budget constraint

$$W_a = \left(1 + \hat{i}(a)\right) W_{a-1} + y_a - c_a \tag{14}$$

We assume that there are no bequest motives such that $W_A = 0$. The Lagrange function of the optimization problem with the new constraint becomes:

$$\mathscr{L} = \sum_{a=0}^{T} \left(\frac{\left(\gamma_a\left(r;k,\alpha\right) c_a\right)^{1-\rho}}{1-\rho} \beta_a - \lambda_a \left(W_a - (1+\psi_a(W_{a-1}))W_{a-1} - y_a + c_a\right) \right) \tag{15}$$

Deriving and solving the first order conditions of the Lagrange function we get:

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial c_a} &= \gamma_a\left(r;k,\alpha\right)^{1-\rho} c_a^{-\rho} \beta_a - \lambda_a = 0 \\ &\Leftrightarrow \gamma_a\left(r;k,\alpha\right)^{1-\rho} c_a^{-\rho} \beta_a = \lambda_a \end{aligned} \tag{16}$$

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial W_a} &= \begin{cases} -\lambda_a W_a + \lambda_{a+1}(1+i_d(a))W_a & \text{if } W_a < 0 \\ -\lambda_a W_a + \lambda_{a+1}(1+i_b)W_a & \text{if } W_a \geq 0 \end{cases} = 0 \\ &\Leftrightarrow \left(-\lambda_a + \lambda_{a+1}(1+\hat{i}(a+1))\right) W_a = 0 \end{aligned} \tag{17}$$

In order for equation 17 to apply for all values of W_a we must have

$$\lambda_a = \lambda_{a+1}(1+\hat{i}(a+1)) \tag{18}$$

Recall that $\beta_a \equiv \prod_{a=1}^{a} \frac{1-\mu_a}{1+\theta}$ such that $\frac{\beta_{a+1}}{\beta_a} = \frac{1-\mu_{a+1}}{1+\theta}$. Equation (16) and (18) combines to the Euler equation:

$$\begin{aligned} &\gamma_a\left(r;k,\alpha\right)^{1-\rho} c_a^{-\rho} \beta_a = \gamma_{a+1}\left(r;k,\alpha\right)^{1-\rho} c_{a+1}^{-\rho} \beta_{a+1}(1+\hat{i}(a+1)) \\ &\Leftrightarrow \gamma_a\left(r;k,\alpha\right)^{1-\rho} c_a^{-\rho} = \gamma_{a+1}\left(r;k,\alpha\right)^{1-\rho} c_{a+1}^{-\rho} \frac{1-\mu_{a+1}}{1+\theta}(1+\hat{i}(a+1)) \\ &\Leftrightarrow c_a = \left(\frac{\gamma_{a+1}\left(r;k,\alpha\right)}{\gamma_a\left(r;k,\alpha\right)}\right)^{\frac{\rho-1}{\rho}} c_{a+1} \left(\frac{1-\mu_{a+1}}{1+\theta}(1+\hat{i}(a+1))\right)^{-\frac{1}{\rho}} \end{aligned} \tag{19}$$

6.2 Endogenous Grid Method

In order to solve the consumer's problem described in previous section, the Endogenous Grid Method (EGM) proposed by Carroll (2006) is applied. Carroll developed the method to solve stochastic optimization problems. Due to the perfect foresight assumption, the applied model assumes uncertainty only with respect to time of death.

As death is an exit state, our optimization problem is of deterministic character, why a more simple version of the EGM method is used. The following algorithm follows the notation and structure of the one in Stephensen (2009), only without stochastics.

First, define a wealth grid with J^W grid points - a grid point is given by $j \in \left\{1,...,J^W\right\}$. The wealth in a given grid point is given by $W(j)$ and this value is given by the definition of the given wealth grid. The defined wealth grid reaches from $W(1) = W_{min}$ to $W(J^W) = Wmax$ and is ordered such that $W(j) < W(j+1)$ for all $j < J^W$. The wealth grid does not solve the optimization problem itself, but it provides the framing of the optimization problem - an invisible, underlying structure determining the span of the optimization procedure. Given a person's initial wealth and future income sequence, all of his future wealth positions should be within the scope of the defined wealth grid. As we allow for negative wealth (debt) we have $W(1) = W_{min} < 0$.

In the following notation, bars denote points that are not situated on the wealth grid. We define a policy function ξ that, for a given age and previous-period net wealth, defines the optimal consumption: $c_a = \xi\left(W_{a-1}, a\right)$. We define the policy function *on the wealth grid* as:

$$c_a(j) = \xi\left(W(j), a\right) = \xi_a(j)$$

From the wealth grid we derive a policy grid which, for each point on the wealth grid, computes the corresponding optimal consumption. So we are actually dealing with two separate grids, a wealth and a policy grid, which relate as in the above equation. We assume that the consumers don't wish to leave any wealth behind why we must have $W_A = 0$. Given this last-period restriction we are able to calculate the last period consumption:

$$c_A = \xi\left(W_{A-1}, A\right) = \left(1 + \psi_A(W_{A-1})\right) W_{A-1} + y_A$$

Or when evaluated on the wealth grid:

$$c_A(j) = \xi_A(j) = \left(1 + \psi_A\left(W(j)\right)\right) W(j) + y_A$$

As the policy function at the terminal age A is known we are able to derive the previous policy functions recursively: Assuming that ξ_{a+1} is known, we want to calculate ξ_a. To do this we use the Euler equation in (19):

$$c_a = \left(\frac{\gamma_{a+1}\left(r;k,\alpha\right)}{\gamma_a\left(r;k,\alpha\right)}\right)^{\frac{\rho-1}{\rho}} c_{a+1}\left(\frac{1-\mu_{a+1}}{1+\theta}(1+\hat{i}(a+1))\right)^{-\frac{1}{\rho}}$$

If we define $\overline{c}_a(j) \equiv \xi_a(j)$ such that $\overline{c}_a(j)$ denotes the level of consumption which directly corresponds to the previous periods' end-of-period wealth being $W(j)$. The above Euler equation evaluated on the grid becomes:

$$\overline{c}_a(j) = \left(\frac{\gamma_{a+1}\left(r;k,\alpha\right)}{\gamma_a\left(r;k,\alpha\right)}\right)^{\frac{\rho-1}{\rho}} \xi_{a+1}(j)\left(\frac{1-\mu_{a+1}}{1+\theta}(1+\hat{i}(a+1))\right)^{-\frac{1}{\rho}} \tag{20}$$

For a given wealth in period $a+1$, equation 20 returns the period a consumption. Let $\overline{W}_{a-1}(j)$ denote the period $a-1$ wealth corresponding to the period a consumption level $\overline{c}_a(j)$ and period a wealth $W_a(j)$. Inserting this in the budget restriction defined in (14) we get

$$W(j) = \left(1+\psi_a(\overline{W}_{a-1}(j))\right)\overline{W}_{a-1}(j) + y_a - \overline{c}_a(j) \Rightarrow$$
$$\overline{W}_{a-1}(j) = \frac{W(j)+\overline{c}_a(j)-y_a}{1+\psi_a\left(\overline{W}_{a-1}(j)\right)}$$

We notice that $\overline{W}_{a-1}(j)$ appears on both sides of the equality. But this is not a problem since the function $\psi_a\left(\overline{W}_{a-1}(j)\right)$ only requires information on the sign of $\overline{W}_{a-1}$ which equals the sign of the numerator $W(j)+\overline{c}_a(j)-y_a$ as the denominator $1+\psi_a\left(\overline{W}_{a-1}\right)$ is always positive. The above equation is rewritten:

$$\overline{W}_{a-1}(j) = \begin{cases} \frac{W(j)+\overline{c}_a(j)-y_a}{1+i_d(a)} & \text{if } W(j)+\overline{c}_a(j)-y_a<0 \\ \frac{W(j)+\overline{c}_a(j)-y_a}{1+i_b} & \text{else} \end{cases}$$

We found the previous-period wealth ($\overline{W}_{a-1}(j)$) and the same-period level of consumption ($\overline{c}_a(j)$) corresponding to a point on the wealth grid $W(j)$. But we are not done yet: $\overline{W}_{a-1}(j)$ is not situated on the wealth grid. In order to proceed to period $a-2$ and derive ψ_{a-1} from ψ_a, $W_{a-1}(j)$ must be situated on the wealth grid. The policy function must be approximated to some value that ensures that $W_{a-1}(j)$ is situated on the wealth grid. This is done by linear interpolation of $\overline{c}_a(j)$.

Assume that $\overline{W}_{a-1}(j)$ is evaluated for all grid points. For every grid point $i \in \left\{1,...,J^W\right\}$, identify j such that $\overline{W}_{a-1}(j) < W(i) < \overline{W}_{a-1}(j+1)$. Now we have that

$$\xi_a(i) = \overline{c}_a(j) + \frac{W(i)-\overline{W}_{a-1}(j)}{\overline{W}_{a-1}(j+1)-\overline{W}_{a-1}(j)}\left(\overline{c}_a(j+1)-\overline{c}_a(j)\right)$$

We don't actually compute $W_{a-1}(j)$ as we only need $c_a(j)$ in order to derive $c_{a-1}(j)$ and so forth. The process of deriving ψ_a from ψ_{a+1} is done in the below steps:

$$\psi_{a+1} = c_{a+1} \Rightarrow \overline{c}_a \Rightarrow \overline{W}_{a-1}(j) \Rightarrow \psi_a \tag{21}$$

The sequence of which the entire policy grid is computed in each grid point j is as follows:

$$W_A = 0 \Rightarrow c_A(j) \Rightarrow \overline{c}_{A-1}(j) \Rightarrow \overline{W}_{A-2}(j) \Rightarrow c_{A-1}(j) \Rightarrow \overline{c}_{A-2}(j) \Rightarrow \overline{W}_{A-3}(j) \Rightarrow \cdots$$
$$\Rightarrow \overline{c}_1(j) \Rightarrow \overline{W}_0(j) \Rightarrow c_1(j)$$

Once the policy grid is defined recursively for all periods, we are able to derive the corresponding optimal consumption paths. This is done in chronological order, beginning with $a = 0$. We consider the initial wealth, W_0 and place it on the wealth grid. We find it to be somewhere between grid point j and $j+1$, let's say $W_0 = bW(j) + (1-b)W(j+1)$ for some $b \in [0,1]$. Now the optimal period 1 consumption is given by

$$c_1 = b\hat{\xi}_1(j) + (1-b)\hat{\xi}_1(j+1)$$

This process is repeated until the entire optimal consumption path is found. Knowing the optimal *consumption* in one period enables us to derive the corresponding *wealth* in the same period which is given in the budget equation in (14). We place this value of W_a on the wealth grid, find it to lie between grid point j and $j+1$ with distance b and get:

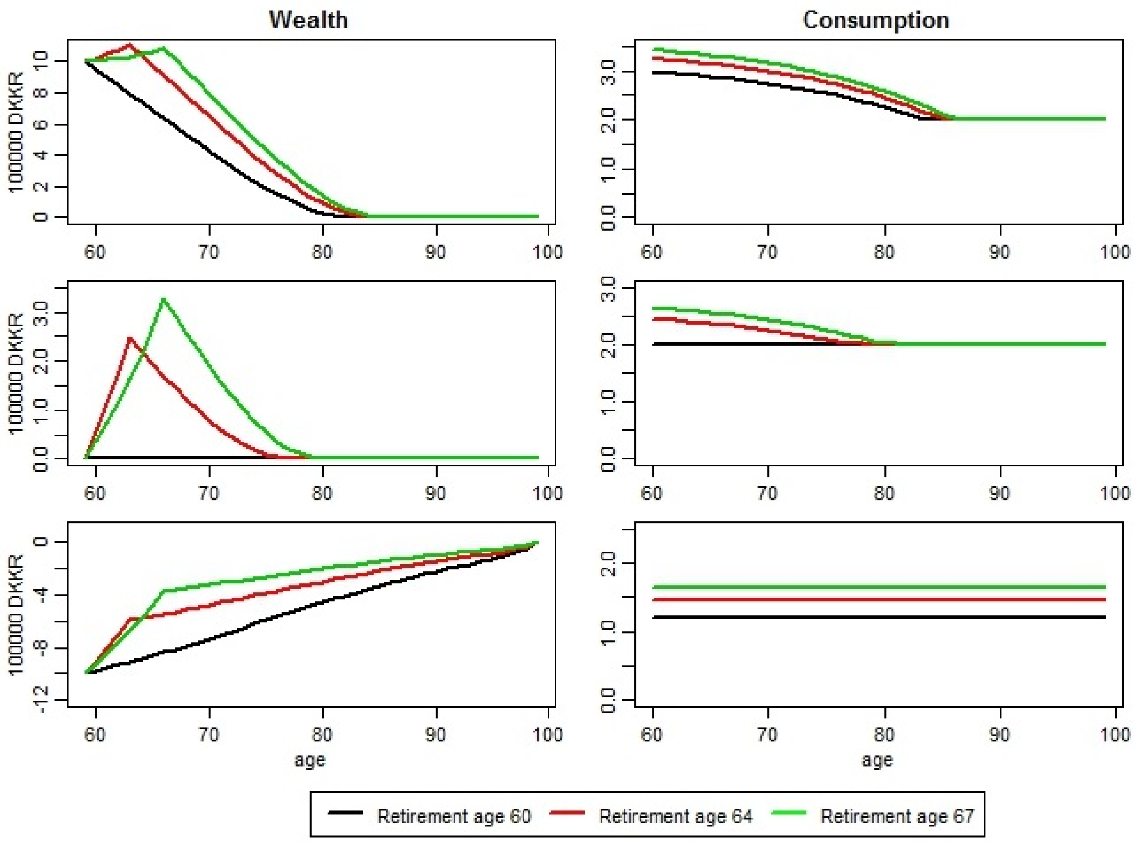
$$c_{a+1} = b\hat{\xi}_a(j) + (1-b)\hat{\xi}_a(j+1)$$

The process is continued until the entire optimal consumption path has been computed.

6.3 Examples of optimal consumption paths

In order to understand what the optimized consumption paths look like we will solve the optimization problem for an individual who earns 300.000 DKK per year prior to retirement and subsequently receives retirement benefits of 200.000 DKK per year. We put $\alpha = 0$ and $k = 1$ (no push effect and no gain/loss in utility level after retirement) and use male death probabilities. Figure 9 show the consumption path and the corresponding development of the individual's wealth when $r = \theta$. The first pair of graphs depicts the solution when the initial wealth is 1 million DKK, the middle pair when it is 0, while the last pair show the solution when negative initial wealth is -1 million DKK. Individuals with positive initial wealth have high initial consumption, higher than their income why they start out to consume all their wealth. As soon as all wealth is consumed, consumption is stabilized at the income level. With no initial wealth accumulation, individuals will only accumulate wealth if they postpone their retirement to benefit from the higher labor market income. Like the case with positive initial wealth, individuals choose a high initial consumption level which decreases until there is no mo accumulated wealth left and consumption stabilize at the income level. Individuals with (significant) negative initial wealth will allays be in debt and the resulting consumption level remains constant. As the different colors symbolizes different retirement ages, they also show how much postponing retirement increases the resulting consumption level. The analysis in Arnberg and Stephensen (2015) is

Figure 9: Optimal Consumption Paths when $r = \theta$



founded on the assumption of a perfect capital market, where both debt and wealth have same actuarial fair interest rate $1 + i_r = (1 + i_a)(1 - \mu_a)$. When $r = \theta$, the resulting optimal consumption path is constant regardless of initial wealth, corresponding to the above case with negative initial wealth where the accumulated wealth never reaches a positive value.

The kink in the consumption path is caused by an asymmetry of the interest function $\hat{i}(a)$. Recall Euler's equation given in equation 19. Inserting the interest rate $\hat{i}(a+1)$ defined in equation 12, $k = 1$, $\alpha = 0$ and $\rho = 2$ we get two different consumption-developments, one characterizing the "wealth-regime" applicable when $W_a > 0$ and one characterizing the "debt-regime" applicable when $W_a \leq 0$:

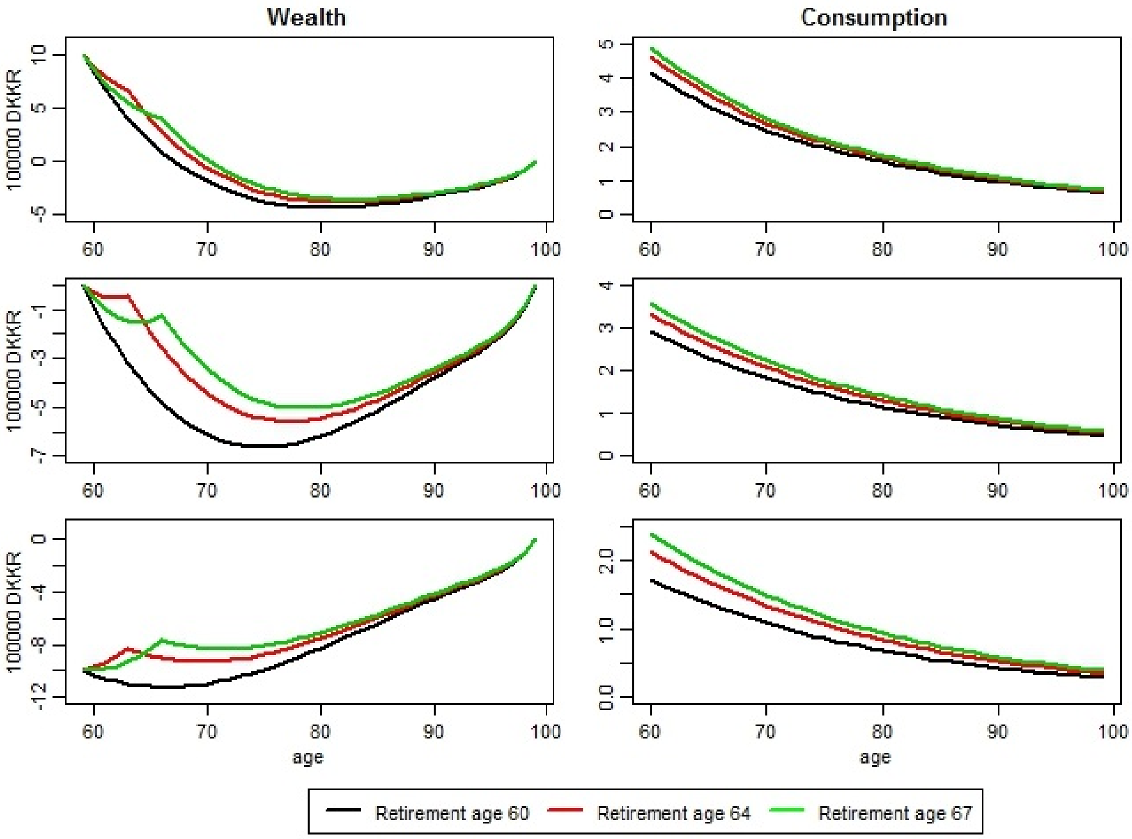
$$c_a = \begin{cases} c_{a+1}\left(\frac{1+i_b}{1+\theta}\right)^{-\frac{1}{2}} & \text{if } W_a < 0 \\ c_{a+1}\left(\frac{1-\mu_{a+1}}{1+\theta}(1+i_b)\right)^{-\frac{1}{2}} & \text{if } W_a \geq 0 \end{cases}$$

As long as wealth is positive, we have that $c_a > c_{a+1}$ and as the death probability increases with a the difference between c_a and c_{a+1} gets larger resulting in the concave shape of the consumption development for the initial years with positive wealth.

As soon as the accumulated wealth is spent and reaches negative values, the individual enters the debt-regime and starts to follow the rule that $c_a = c_{a+1}$ if $i_b = \theta$. The individual just reach an infinitesimal negative amount of wealth which is kept constant due to this rule where consumption is kept at a steady level equal to the level of income.

However, we usually find that the subjective time preference rate θ is higher than the interest rate [5], reflecting impatience of the consumer. Figure 10 show how the chosen wealth- and consumption paths change when peoples impatience increases.

Figure 10: Optimal Consumption Paths when $r = 4.75\%$ and $\theta = 15\%$



Now, consumption in the beginning of the period is remarkably higher than before and the consumption level decreases throughout the entire period for all three cases of initial wealth. First, individuals dis-save in order to finance the high initial consumption levels. The consumption level decreases, and at some point it gets below the income level, why individuals again start to accumulate wealth such that at the terminal age, $A = 100$, they have repaid all their accumulated debts. The consumers

[5]Arnberg and Stephensen (2015) advocate for θ parameters around 0.1-0.2, referring to Andersen et al. (2008) who bases their analysis on an experimental design with approximately 250 Danish participants

actually chose borrow money with guarantee in their future retirement benefits when θ is large enough.

A high level of impatience leads to a decreasing consumption level. If you don't value consumption in the future as much as consumption today there is no reason not to spend more today and compensate with a low future consumption. The higher θ, the steeper consumption path, also reflected in the Euler equation. The fact that $\theta > i_b$ results in a decreasing consumption paths in both the wealth- and debt regime. In the upper example with positive initial wealth, the decrease in consumption is slightly steeper during the wealth-regime, but due to the relative low death probabilities prior to age 70 the difference is very small and impossible to detect in the graphs.

7 Estimation

The model allows for heterogeneous preferences for leisure, k, and a non-parametric estimation technique is developed to measure this heterogeneity. The following show how, with an iterative process, one can make the average distribution of preferences converge towards the true population distribution. The basic idea behind this estimation technique resembles that in Train (2007) but with significant differences: Whereas Train (2007) identifies the moments of some theoretical distribution, the current method is non-parametric. The theory underlying the estimation in Train (2007) also applies to the non-parametric estimation in the current study and is further elaborated in Train (2009) chapter 11 and 12. The remaining homogeneous model parameters, α, ϕ and θ are measured with Maximum Likelihood Estimation.

7.1 Estimation of *k*

k is assumed to be a heterogeneous parameter meaning that it can vary for different individuals in the population. One of the primary goals of this study is to estimate the population distribution of k non-parametrically. The following section explains how this distribution is derived from the combination of observed and simulated data described in the previous section

Let the population consist of N individuals. For each individual $j, j \in \{1,...,n\}$ the initial wealth W_0^j is observed together with the actual retirement age, r^j. Let $Y(r)^j$ denote the individuals' income sequence given retirement age r, such that

$$Y(r)^j = \left(Y_1^j,...,Y_{r-1}^j,B_r^j,B_{r+1}^j,...,B_A^j\right)$$

with Y_a^j (B_a^j) denoting the individuals' income prior to (after) retirement. From observed data on salaries, pension savings etc. together with knowledge of the tax and public transfer system, the income sequences $Y_1^j,...,Y_P^j$ are simulated as proposed in section 5.3.

Given the logit specification denoted in equation 8 and derived in appendix 10.1, the probability of individual j retiring at age r is given by

$$P^j\left(r|k^j,\alpha,\phi,\theta\right) = \frac{exp\left(\phi V_0^j(r)\right)}{\sum_{a=1}^P exp\left(\phi V_0^j(a)\right)} \tag{22}$$

As stated in equation 6, $V_0(r)$ is defined as the utility of all future consumption at age $a = 0$:

$$V_0^j(r) = \sum_{a=1}^{A} \frac{\left(\gamma_a\left(r;k^j,\alpha\right)c_a^j\right)^{1-\rho}}{1-\rho}\beta_a \tag{23}$$

The future consumption path of individual j can be derived numerically from $x^j = \left(W_0^j, Y(1)^j, ..., Y(P)^j\right)$ applying the Endogenous Grid Method (EGM) as described in chapter 6. The probability of retiring at a given age can therefore be considered as a function of x^j alone. This enables us to express the probability of person j's retirement age as a general function, π. This function applies for all individuals but requires an individual specific input x^j:

$$\pi\left(r|k,\alpha,\phi,\theta,x^j\right) = \frac{exp\left(\phi V_0^j(r)\right)}{\sum_{a=1}^{P} exp\left(\phi V_0^j(a)\right)} \tag{24}$$

Equivalently to the assumption that x^j is drawn from some distribution with density function $f(x), x \in X$ we assume that k_j is drawn from a distribution with density function $p(k)$, $k \in \mathcal{K}$. Doing this we assume the distributions of x^j and k to be independent which we will return to in 7.5.

We observe/simulate x^j while k^j is non-observable to the researcher. We will now show how the distribution $p(k)$ can be derived by non-parametric estimation given the data d^j for $j \in \{1,...,N\}$. We begin with Bayes' rule, which relates the conditional to prior believe of two events, say event A and event B. We have that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

According to Bayes' rule we are able to exploit the information about a persons actual retirement age a_{r_0} to say something about the probability of that person's k-parameter. The underlying intuition is simple: A person retiring at a late age indicates a low value of k, while an early retirement indicates a high value. Applying Bayes' rule to r and k gives us:

$$P^j(k|r^j;\alpha,\phi,\theta) = \frac{P^j(r^j|k,\alpha,\phi,\theta)p(k)}{P(r^j;\alpha,\phi,\theta)} = \frac{\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)p(k)}{P(r^j;\alpha,\phi,\theta)} \tag{25}$$

Here $p(k)$ denotes the population distribution and $P^j(k|r^j;\alpha,\phi,\theta)$ the conditional distribution of k as a function of α, ϕ and θ.

Now we are able to derive the unconditional probability of r^j from the its conditional distribution $\pi\left(r|k,\alpha,\phi,\theta,x^j\right)$ defined in equation (24). By marginalizing out the conditional variable we obtain the unconditional probability:

$$P(r^j;\alpha,\phi,\theta) = \int_{\mathcal{K}} \pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)p(k)dk$$

However, we still have no clue as to what the item of interest, $p(k)$, is. But even though $p(k)$ appears on the right hand side of (25), it can still help us find $p(k)$ iteratively. By $P(r^j;\alpha,\phi,\theta)$ we've found a way to describe each individuals' k-distribution.

And if we consider the average of all individual's $P(r^j;\alpha,\phi,\theta)$ we are able to describe k's distribution in the entire population. The following result is indeed crucial to the analysis.

Theorem 7.1.1 (Estimation of the k-distribution)**.** *Consider N consumers and let the probability of consumer j retiring at age a be given by* (22)*. Assume that x_j is drawn from a distribution with density function $f(x)$ and k from a distribution with density function $p(k)$. Let $P^j(k|r^j)$ be given by* (25)*. If and only if $P^j(k|rj)$ is derived from the correct prior $p(k)$ we have:*

$$p(k) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} P^j(k|r^j)$$

Proof. Define

$$\hat{P}^j(k|r^j) = \frac{\pi\left(r|k,\alpha,\phi,\theta,x^j\right)\hat{p}(k)}{\hat{P}(r;\alpha,\phi,\theta)} \tag{26}$$

Where

$$\hat{P}(r;\alpha,\phi,\theta) = \int_0^\infty \pi\left(a^j|k,\alpha,\phi,\theta,x^j\right)\hat{p}(k)dk$$

The "hat" in $\hat{p}(k)$ indicates, that we are not necessarily referring to the true prior $p(k)$, but some guess. Now we have that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{N}\sum_{j=1}^{N}\hat{P}^j(k|a^j) &= E\left[\hat{P}^j(k|a^j)\right] \\
&= E\left[\frac{\pi\left(r|k,\alpha,\phi,\theta,x^j\right)\hat{p}(k)}{\hat{P}(r;\alpha,\phi,\theta)}\right] \\
&= \int_X \sum_{a_r=1}^{P} \frac{\pi\left(r|k,\alpha,\phi,\theta,x^j\right)}{\hat{P}(r;\alpha,\phi,\theta,x)}\hat{p}(k)P(r;\phi,x)f(x)dx \\
&= \hat{p}(k)\int_X \sum_{a_r=1}^{P} \pi\left(r|k,\alpha,\phi,\theta,x^j\right)\frac{P(r;\alpha,\phi,\theta,x)}{\hat{P}(r;\alpha,\phi,\theta,x)}f(x)dx
\end{aligned}$$

Only if the guess $\hat{p}(k)$ is correct we have that $\hat{p}(k) = p(k)$ and thereby $\forall r \in \{1,...,P\}$, $x \in X$ we have $P(r;\alpha,\phi,\theta,x) = \hat{P}(r;\alpha,\phi,\theta,x)$ such that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{N}\sum_{j=1}^{N}P^j(k|a^j) &= p(k)\int_X \sum_{a_r=1}^{P} \pi\left(r|k,\alpha,\phi,\theta,x^j\right)\frac{P(r;\alpha,\phi,\theta,x)}{P(r;\alpha,\phi,\theta,x)}f(x)dx \\
&= p(k)\int_X \sum_{a_r=1}^{P} \pi\left(r|k,\alpha,\phi,\theta,x^j\right)f(x)dx \\
&= p(k)\int_X f(x)dx \\
&= p(k)
\end{aligned}$$

□

Theorem 7.1.1 looks promising. It states that the mean of all individuals' k-probabilities conditioned on their observables r^j and x^j approaches $p(k)$ when the number of individuals goes to infinity. Naturally we don't have infinitely many individuals in our population (and certainly not in our data), but we will treat the asymptotic results as approximately valid for our finite sample sizes. However, we still have one crucial problem: in order to compute the conditional k-probabilities in (25), we must know the unconditional population probability $p(k)$. Inserting equation (25) in theorem 7.1.1 we get

$$p(k) = \frac{1}{N} \sum_{j=1}^{N} \frac{\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)p(k)}{\int_{\mathcal{K}} \pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)p(k)dk} \tag{27}$$

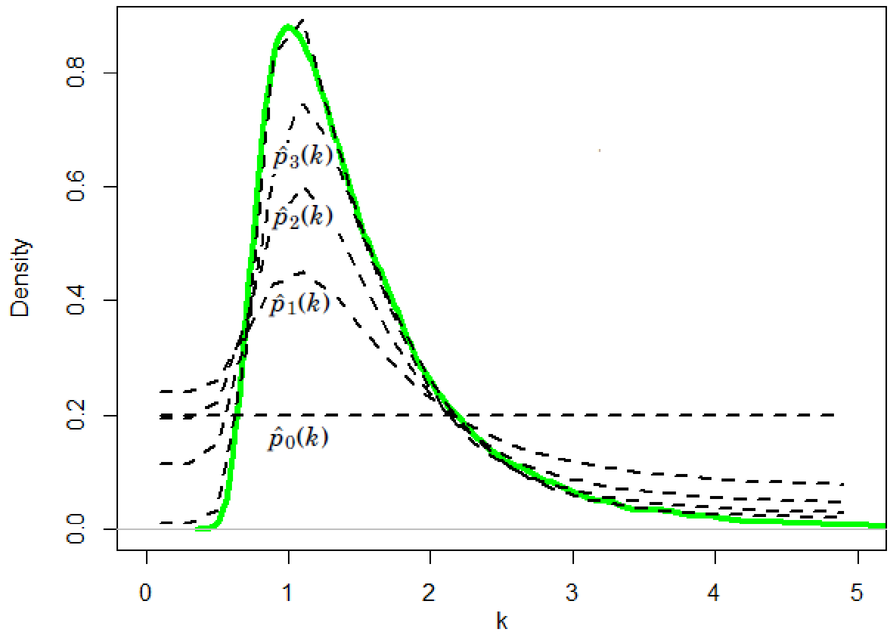
What we have is a fixed-point problem: We want to find the distribution $p(k)$ that solves $p(k) = f(p(k))$, f being the function in above equation 27. It turns out that if we just insert some arbitrary, random guess $\hat{p}(k)$ in the right hand side of equation 27, compute the left hand side and repeat the computations, we will reach the true distribution of $p(k)$. The reason for this is given in Banach's Fixed Point Theorem, which is stated in appendix 10.2. The explanation of why equation (25) converges is that the function is a contraction mapping. As stated in the theorem in appendix 10.2, we let the distribution $\hat{p}_0(k)$ be some initial guess, e.g. a uniform distribution, and then $\hat{p}_s(k) \to p(k)$ for $s \to \infty$ with

$$\hat{p}_s(k) = \frac{1}{N} \sum_{j=1}^{N} \frac{\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)\hat{p}_{s-1}(k)}{\int_{\mathcal{K}} \pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)\hat{p}_{s-1}(k)dk} \tag{28}$$

Proving that our function is in fact a contraction mapping is highly complex as we are not dealing with the transformation of a value, but an entire distribution. An actual mathematical proof has therefore been left out of the analysis. A more intuitive explanation as to why equation 27 converges to the true prior lies in the Bayes' rule stated in equation 25. Given an individual's information (x^j and r^j), his actual retirement age is more probable fore some values of k than others, which will be reflected in the values of $\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)$ evaluated at different ks. This probability is weighted by the prior probability, $\hat{p}(k)$, but the resulting distribution $P^j(k|r^j;\alpha,\phi,\theta)$ is pushed towards the distribution of $\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)$. The average of all individuals' conditional distributions of k results in the distribution $\hat{p}_s$, which will necessarily be closer to the true distribution than $\hat{p}_{s-1}$ due to the fact that all individuals' $\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)$ has pushed it in a direction that fits data better.

When applying the described estimation theory on simulated data, we see that the estimated distribution does in fact converge to the true distribution. This is illustrated in figure 11 that show how, for each iteration i, the initial uniform distribution converges to the true simulated distribution of k plotted in green. When initiating the estimation with different shapes of initial distributions, they all converge to the true distribution.

Figure 11: Convergence of $\hat{p}(k)$



Note: Based on simulated data. Illustration of how a uniform initial guess $\hat{p}_0(k)$ converges iteratively to the true distribution $p(k)$ (the green curve).

Train (2009) section 11.5 mentions the same convergence property as that in Theorem 7.1.1: for a correctly specified model at the true population parameters, the conditional distribution of tastes, aggregated over all individuals, equals the population distribution of tastes. They refer to Allenby and Rossi (1999) who also applies this property when estimating heterogeneity, however with normality constraints.

7.1.1 Outline of Estimation Algorithm

The following "cook-book" description of the estimation of the k-distribution clarifies how the fixed point problem can be approached:

1. Make a reasonable guess about the boundaries of k, k_{min} and k_{max}.
2. Assume some random distribution of k, $\hat{p}(k)$, e.g. a uniform distribution on the assumed interval $[k_{min}, k_{max}]$.
3. For each individual, derive $\hat{P}^j(k|r)$ from (26). Despite the fact that $\hat{p}(k) \neq p(k)$, $\hat{P}^j(k|r)$ will be affected by x_j and thereby shaped in a way that fits the data.

4. Now let $\hat{p}(k)$ equal the average of $\hat{P}^j(k|r)$.

5. Repeat step 2-4 until convergence of $\hat{p}(k)$.

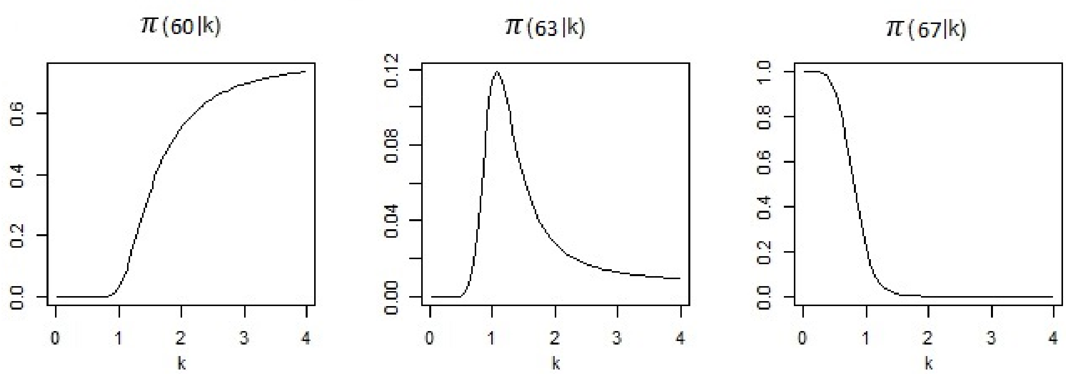
The estimation of $p(k)$ depends on decisions made in the numerical computation. In order to evaluate the continuous distribution of k we need to define a grid, reaching from k_{min} to k_{max} with a given step size. Of course both step size as well as convergence criteria has an effect of the final estimated distribution, why we must be careful to set each small enough as to ensure that they don't bias the estimation results.

7.1.2 Properties of π

$\pi\left(r|k\right)$ denotes the probability of retirement at age r for a given value of k. This section examines the properties of π for different values of r and k, taking the remaining parameters ϕ α and x^j for given.

Letting $\{1,...,P\} = \{60,...,67\}$, figure 12 illustrates π evaluated for retirement at age $1 = 60$, $4 = 63$ and $P = 67$ for $k \in [0.1, 3.9]$ for simulated data.

Figure 12: Graphs depicting $\pi(r|k)$ for $r \in \{1 = 60, 4 = 63, P = 67\}$



Note: Based on simulated data. For the first (last) possible retirement age, π is maximized in k_{max} (k_{min}), increasing (decreasing) as k grows while $\pi(64|k)$ is bell-shaped with a maximum in $k \approx 1.1$

.

Retiring early (r=60) indicates a high k, why $\pi(60|k)$ is highest for large values of k. However, the fact that $\pi(60|k)$ is monotonously increasing in k implies that k is unidentified for an individual who actually retires at age 60. Having $k = 10$ or $k = 100$ makes it equally probable that he retired at age 60. Another issue occurs with late retirements: here, the lowest possible value of k maximizes the probability of retirement at 60. The maximizing value of k is much better defined for intermediate retirement ages.

The fact that k is unidentified for individuals who retire at the earliest possible retirement age $a = 0$ or the latest P has some unfortunate effects on the tails of the estimated distributions. Recall that the measured population distribution $p(k)$ is merely a weighted average of the individuals' π functions. Those individuals retiring at $r = 1$ will contribute with large probabilities for the highest k-values, while those retiring at $r = P$ will contribute with large probabilities for k close to 0. Individuals with exceptional large or small values of k might not be "outweighed" in the averaged distribution as those with late/early retirement contribute with density mass for k_{min} and k_{max}, enabling the estimated distribution to have tails. The tails do not indicate that there are considerable fractions of individuals with extremely large/small k-values but is merely a result of the behavior of π in the limit values of k.

One might think, that as k grows toward infinity, the probability retirement at 1 would approach 1 [6]. This is, however, not the case, and will be proved in the following.

$V_0(r)$ denotes the present-value utility of all future consumption. Its definition given in equation 23 implies that

$$V_0(1) = \sum_{a=1}^{A} \frac{\left(kc_a\right)^{1-\rho}}{1-\rho}\beta_a$$

$$V_0(2) = \frac{\left(e^{-\alpha}c_1\right)^{1-\rho}}{1-\rho}\beta_1 + \sum_{a=2}^{A} \frac{\left(ke^{-\alpha}c_a\right)^{1-\rho}}{1-\rho}\beta_a$$

$$\vdots$$

$$V_0(P) = \sum_{a=1}^{a_{p-1}} \frac{\left(e^{-\alpha(a-0)^2}c_a\right)^{1-\rho}}{1-\rho}\beta_a + \sum_{a=P}^{A} \frac{\left(ke^{-\alpha(P-0)^2}c_a\right)^{1-\rho}}{1-\rho}\beta_a$$

First consider the case where $k \to \infty$

When $\rho = 2$ we get $V_0(a) < 0 \ \forall a \in \{1,...,P\}$ and that $V_0(a) \to 0$ for $k \to \infty$ while $V_0(a) < 0$ for $a \in \{2,...,P\}$, for $k \to \infty$.

Recall the definition of $\pi\left(r|k\right)$ given in equation 24:

$$\pi\left(r|k\right) = \frac{exp\left(\phi V_0(r)\right)}{\sum_{a=1}^{P} exp\left(\phi V_0(r)\right)}$$

Evaluated in $r = 1$, the above expression can be rephrased into:

$$\pi\left(1|k\right) = \frac{exp\left(\phi V_0(1)\right)}{\sum_{a=1}^{P} exp\left(\phi V_0(a)\right)} = \frac{1}{1 + \sum_{a=2}^{P} exp\left(\phi(V_0(a) - V_0(1))\right)}$$

[6] This is falsely stated in Arnberg and Stephensen (2015)

And as $\phi(V_0(a)-V_0(1)) < 0$ for all $a \in \{2,...,P\}$ when $k \to \infty$ it must be that $\pi\left(1|k\right) < 1$ for $k \to \infty$.

The explanation of this contra-intuitive behavior of ϕ lies in the chosen utility specification, more precisely that $\rho = 2$. If $\rho < 1$ then $\pi\left(1|k\right) = 1$ for $k \to \infty$. The CRRA utility-scale with $\rho = 2$ will always be negative (when consumption is positive), and when the level of consumption increases, utility approaches 1. Retirement at age r is chosen as long as

$$U_0(r) > U_0(a) \quad \forall \quad a \neq r \Leftrightarrow$$
$$\phi V_0(r) > \phi V_0(a) + \epsilon_a - \epsilon_r \quad \forall \quad a \neq r$$

The upper limit of V_0 implies that there is a positive probability of an ϵ_a large enough to make an individual prefer retirement at age a, even though consumption at retirement age r is infinitely larger compared to age a. This implies that

$$0 < \lim_{k \to \infty} \pi\left(r|k\right) < 1 \quad \forall \quad \in \{1,...,P\}$$

7.2 Log Likelihood estimation

Estimation of ϕ and α is done by maximizing a log likelihood function. We know that the probability of a person retiring at age r conditional on k is given in equation (22). Given the estimation of the k distribution (see section 7.1) we are able to derive the unconditional probability of a person retiring at age r as a function of x_j, ϕ and α. We derive the unconditional probability from the conditional by marginalizing out the conditional variable k:

$$Pr(r;\alpha,\phi,\theta,x^j) = \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) p(k;\alpha,\phi,\theta)dk$$

Note that we write the prior distribution of k as a function of α and ϕ. This is because the k-distribution is derived on the basis of some specific assumed values of α and ϕ. If we were to change these values, we would arrive at a different measure of $p(k)$. Given that we know the actual retirement ages of the individuals, r^j, the likelihood of our data set given the values of ϕ, α and θ is found as:

$$L(\alpha,\phi,\theta) = \prod_{j=1}^{N} Pr(r^j;\alpha,\phi,\theta,x^j)$$

And hence the log likelihood function becomes

$$LL(\alpha,\phi,\theta) = \sum_{j=1}^{N} log \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) p(k;\alpha,\phi,\theta)dk \tag{29}$$

We wish to find the parameters $\hat{\phi}$ and $\hat{\alpha}$ which minimizes the likelihood function. Due to the numerical derivation of $p(k)$, we need to solve this minimization problem numerically as well. We use the *optim* function in R to search for and find the optimal values of ϕ, α and θ.

7.3 Handling data censoring

Recall our model assumptions about the retirement decision: the retirement age is decided upon at age $a = 0$ where all future financial circumstances are known. All we need in order to compute $x^j = \left(W_0^j, Y(1)^j, ..., Y(P)^j\right)$ is to observe the individual at age $a = 0$. However, in order to observe the retirement age r^j, the individual must be present in the data set when he retires. One could think about presence in our data set as being alive - we will return to this later. The probability of dying at a specific age is given by μ_a.

The probability of individual j retiring at age r, as defined in equation 22, implicitly assumes that individual j is still alive at age r^j. The probability that individual j is still alive at age r^j is the product of the probabilities of the individual living in the years prior to r^j. We will denote this probability by M_r:

$$M_r = \prod_{a=1}^{r} (1 - \mu_a)$$

The probability of retiring at a given age should take this probability into account such that the "mortality corrected" probability, π_μ, becomes

$$\pi_\mu \left(r|k, \alpha, \phi, \theta, x^j\right) = \pi \left(r|k, \alpha, \phi, \theta, x^j\right) M_r$$

How will including mortality effect the estimation process? It turns out it won't. As the death probabilities are constant, they will cancel out throughout the estimations. Including mortality in the estimation of k as given in equation 25 gives:

$$\begin{aligned}
P_\mu^j(k|r^j; \alpha, \phi, \theta) &= \frac{\pi_\mu \left(r^j|k, \alpha, \phi, \theta, x^j\right) p(k)}{\int_{\mathcal{K}} \pi_\mu \left(r^j|k, \alpha, \phi, \theta, x^j\right) p(k) dk} && (30) \\
&= \frac{\pi \left(r^j|k, \alpha, \phi, \theta, x^j\right) p(k) \cdot \cancel{M_r}}{\int_{\mathcal{K}} \pi \left(r^j|k, \alpha, \phi, \theta, x^j\right) p(k) dk \cdot \cancel{M_r}} && (31) \\
&= P^j(k|r^j; \alpha, \phi, \theta) && (32)
\end{aligned}$$

In the log likelihood function given in equation 29 we find that the mortality contribution becomes an additive constant that can be disregarded in the estimation process:

$$
\begin{aligned}
LL^{\mu} &= \sum_{j=1}^{N} \log \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) M_r p(k;\alpha,\phi,\theta) dk &(33)\\
&= \sum_{j=1}^{N} \log \left(\int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) p(k;\alpha,\phi,\theta) dk \cdot M_r \right) &(34)\\
&= \sum_{j=1}^{N} \log \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) p(k;\alpha,\phi,\theta) dk + \sum_{j=1}^{N_l} \log M_r &(35)\\
&= LL + constant &(36)
\end{aligned}
$$

If an individual dies after having retired, nothing changes. But if an individual dies prior to his retirement, things becomes different as we don't observe his retirement age. Now we are no longer interested in the probability of retiring at a given age, but rather the probability of *not retiring prior to dying* which we will denote by π_d.

The probability of dying at age a_d is given by the death probability at a_d times the probability of surviving prior to a_d:

$$
M^d_{a_d} = \mu_{a_d} \prod_{a=1}^{a_d-1} (1-\mu_a)
$$

The probability of not retiring prior to dying at age a_d is the probability of *not retiring* prior to age a_d times the probability of dying at age a_d, $M^d_{a_d}$:

$$
\pi_d\left(r|k,\alpha,\phi,\theta,x^j\right) = \left(1 - \sum_{a=a1}^{a_d} \pi\left(a|k,\alpha,\phi,\theta,x^j\right)\right) M^d_{a_d}
$$

As in the previous case the mortality term will not affect the estimation. The relevant function for estimating k becomes

$$
P^j_d(k|a^j_d;\alpha,\phi,\theta) = \frac{\left(1-\sum_{a=a1}^{a_d} \pi\left(a|k,\alpha,\phi,\theta,x^j\right)\right) p(k) \cdot \cancel{M^d_{a_d}}}{\int_{\mathcal{K}} \left(1-\sum_{a=a1}^{a_d} \pi\left(a|k,\alpha,\phi,\theta,x^j\right)\right) p(k) dk \cdot \cancel{M^d_{a_d}}}
$$

And the log likelihood

$$
LL^d = \sum_{j=1}^{N_d} \log \int_{\mathcal{K}} \left(1 - \sum_{a=a1}^{a_d} \pi\left(a|k,\alpha,\phi,\theta,x^j\right)\right) p(k;\alpha,\phi,\theta) dk + \sum_{j=1}^{d} \log M^d_{a_d}
$$

Now we have defined the different estimation techniques for the two groups. We need to include both techniques simultaneously in order to estimate on the total sample including those with and without observed retirement age. Define

$$
\bar{\pi}\left(r|k,\alpha,\phi,\theta,x^j\right) = \begin{cases} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) & \text{if } a_d > r \\ \left(1-\sum_{a=a1}^{a_d} \pi\left(a|k,\alpha,\phi,\theta,x^j\right)\right) & \text{if } a_d \le r \end{cases}
$$

Replacing π with $\bar{\pi}$ in the estimation of k and the log likelihood function enables us to include individuals without observed retirement ages and thereby avoid potential selection bias. Of course, not all individuals disappear in data due to death. Some might emigrate and others might disappear simply due to some error in the data collection. However, as the probabilities of dying (or more generally the probability to disappear in data) either cancel outs or enters the log likelihood function as some constant term, they can simply be disregarded.

7.4 Structure of the estimation program

What we are dealing with is a nested fixed point algorithm. The estimation of α and ϕ is done by maximizing a likelihood function. But nested in this likelihood function lies both the EGM-function generating the optimal consumption path and the iterative estimation of the k-distribution as described in 7.1. This makes the estimation procedure quite complex and an analytical solution impossible. The maximization of the log likelihood must be done by searching through different combinations of α and ϕ until the likelihood function is maximized. For each guess of $\hat{\alpha}$ and $\hat{\phi}$, the corresponding consumption paths and k-distribution must be computed. The steps in the estimation process can be put as follows:

(1) Assume values of $\hat{\alpha}$, $\hat{\phi}$ and $\hat{\theta}$.

   (i) For all k-values in the grid $k \in \{k_{min}, ..., k_{max}\}$

      (A) For all individuals $j \in \{1, ..., N\}$

         (i) For all possible retirement ages $a \in \{1, ..., P\}$

            (a) Derive optimal consumption path with EGM

            (b) Derive utility

      (B) Derive $\pi\left(r^j | k, \alpha, \phi, \theta, x^j\right)$ for all k-values, r^j being the actual retirement age of individual j

   (ii) Find P_j for all k-values in the grid and update k-distribution. Until convergence return to (i)

(2) When the k-distribution is converged for the given $\hat{\alpha}$, $\hat{\phi}$ and $\hat{\theta}$, derive log likelihood value.

(3) Return to (1) with new values of $\hat{\alpha}$, $\hat{\phi}$ and $\hat{\theta}$ and continue until the maximum likelihood is found

7.5 Critical assumptions

During the previous derivation of the model and estimation procedures we made several assumptions with the purpose to simplify and enable estimation. It is of highest importance that the researcher considers the underlying assumptions of his model: if they are likely to bias the estimation results, if they could be avoided or if their presence affects the interpretation of the estimation results. In this section we will discuss three of the more critical assumptions made: uncorrelated error terms, independence of k and x and perfect foresight.

7.5.1 Uncorrelated error terms

In section 3.3 we defined utility to consist of a deterministic part, V_0 and a random part ϵ which is known only to the individual. Our logit model relies on the assumption that the error terms are uncorrelated. The assumption is equivalent to assuming that $V_0(r)$ is sufficiently well specified such that the remaining unobserved utility is essentially "white noise". Our specification of $V_0(r)$, the representative utility, takes account of the optimal consumption path for an individual given his future income and initial wealth. It also includes the heterogeneous parameter k and the attrition parameter α. However, the remaining contribution to the utility lies in the error terms ϵ. It seems reasonable to assume, that if a person's ϵ is large for one retirement age, he's likely to have a positive epsilon for the following retirement age - correlation in the error terms seems likely.

The behavioral implication of IID error terms is independence of irrelevant alternatives, the IIA property. The IIA property implies that the probability ratio of individuals choosing between two retirement ages, i and j, does not depend on the availability or attributes of the other retirement ages:

$$\frac{Pr\left(i|k,\alpha,\phi,\theta\right)}{Pr\left(j|k,\alpha,\phi,\theta\right)} = \frac{\exp\left(\phi V_0(i)\right)/\sum_{a=0}^{P}\exp\left(\phi V_0(a)\right)}{\exp\left(\phi V_0(k)\right)/\sum_{a=0}^{P}\exp\left(\phi V_0(a)\right)} = \exp\left(\phi\left(V_0(i) - V_0(j)\right)\right)$$

If the IIA property is plausible, estimation on any subset of alternatives would result in parameter estimates not significantly different from those on the full set of alternatives as described in Hausman and McFadden (1984). The IIA property implies independent error terms, why a test of the IIA property could be applied to test the plausibility of the IID assumption. Due to time limitations such a test is not performed in the current analysis. Several models relax the IIA assumption by allowing for more flexible correlation structures in the error terms. The Multinomial Probit model is build on the assumption that the error terms are multivariate normal with arbitrary correlations between ϵ_i and ϵ^j for all $i \neq j$. Another popular alternative is the Nested

Logit model, which parts the set of alternatives into subsets called "nests". The IIA property is assumed to hold inside the nests but not across different nests. The flexibility of the Multinomial Probit and the nested Logit model, however, comes at a price as the estimation is numerically very demanding.

7.5.2 Independence of k and x

The nonparametric estimation of the k-distribution relies on the assumption that k and x are independent. Recall that $x = \left(W_0, Y(0), ..., Y(P)\right)$ covers initial wealth and income streams for all retirement ages. One theory could be that individuals with low (high) k-values will tend to have high (low) combination of income and wealth. The explanation being that high-wage jobs are more meaningful, interesting and forging identity than low wage jobs, why high-earners value retirement less compared to working than low-earners, resulting in smaller k's. Some degree of dependence between k and x seems fairly reasonable.

How a violation of the assumption affect estimation can be investigated through data simulation.

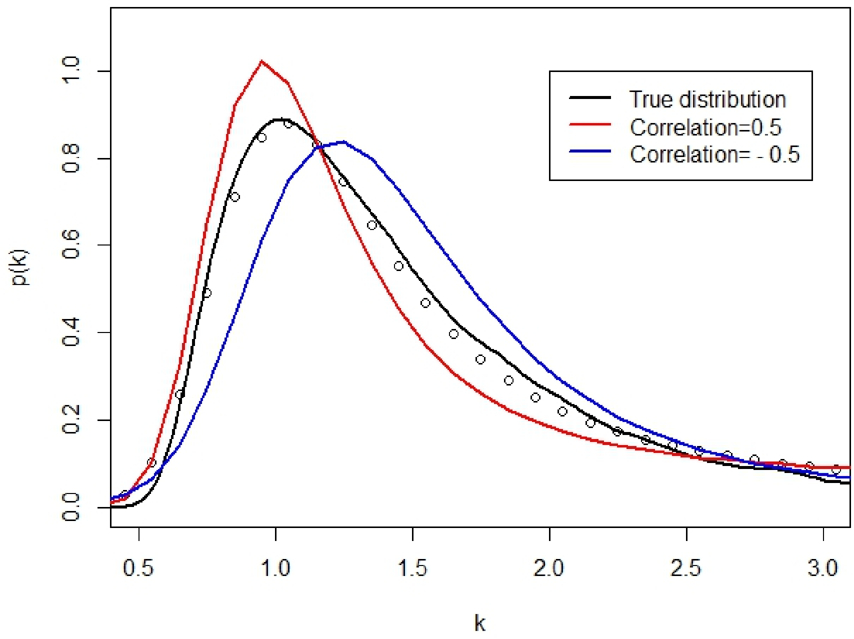
We simulate data for given values of α, ϕ and θ, draw income and initial wealth from log-normal distributions and define a simple retirement system. We draw different samples of k-values, all with same distribution but different correlations with x. Estimating the k-distribution from data simulated from the different k-simulations show how a correlation affects the estimated k-distribution.

We simulate 50.000 individuals and draw k, W and Y (in 100.000s) from log-normal distributions. The data is simulated from our model with $\alpha = 0.005$, $\phi = 2$, $\Theta = 0.015$ and $\rho = 2$. Figure 13 show the estimation bias that occurs at different degrees of correlation and that a higher correlation leads to more bias. The solid curves depicts the true density function of our simulated k's and the dots the estimated density in the given grid point. We are able to estimate the $k$distribution correctly when k and x are uncorrelated but a correlation of 0.5 (-0.5) skews the estimated curve to the right (left) compared to the true distribution. A correlation of ± 0.5 is quite high, and while the estimated distributions are clearly biased, the degree of bias is not overwhelming. What we can do to minimize the correlation in the data is to divide the data-set into subgroups that have a smaller variation in x. A smaller variation in x will make the bias less prevalent. We divide the data by gender and 6 education levels.

7.5.3 Perfect foresight

Assuming that individuals have perfect foresight poses some challenges. Perfect foresight with respect to future income requires knowledge of unpredicTable matters, for example future benefit rates, job situation, health and partner status. In reality, final

Figure 13: Correlation Bias of k-distribution



Note: The black curve is the true simulated density of k. The dotted line is the estimation when there is no correlation between k and x^j. When we simulate k such that the k-distribution remains the same but with a correlation of 0.5 with x^j (red curve) we estimate a distribution which is skewed to the right compared to the true distribution. If the correlation is -0.5 (blue curve) the estimated distribution is skewed to the left

.

retirement decisions are not taken at age $a = 0$. One might decide when to retire at age $a = 0$, but this decision might change as new information is available, e.g. changes in health, job status or even the tax- or benefit system.

Having said that, it is reasonable to assume that the perfect foresight assumption is more appropriate in a Danish context compared to e.g. a US context. As the Danish welfare system is universal and free health care is available to all, health shocks are less significant. The Danish welfare system also constitutes a strong safety net such that individuals who lose their job or working ability are provided relatively generous social assistance, unemployment- and disability benefits.

When computing the future income streams we should, strictly speaking, not include any changes that were unpredicTable at age 59. In accordance with the model assumption, we should base our computations based on the exact situation at age $a = 0$ and then extrapolate everything with some growth and inflation rate. But being 100% true to our model assumptions on perfect foresight might make our calculations unnecessary unrealistic and give inaccurate pictures of the actual situations the individual faces. An individual who hasn't retired, but receives new information after age $a = 0$, would include this information in his decision making.

When computing the future income streams we chose to contradict the assumed retirement decision process and include events that were not known to the individual at age $a = 0$, but which occurred prior to the latest possible retirement age, P.

One example is the rule of increased OAP that was introduced in 2004: postponing retirement 1 year after age 65 would result in a 6% increase in all future OAP benefits, 2 year in a 12% increase. While an individual born in 1941 would not know of this reform in 2001, he would know of it at age 63 in 2004. Given that he didn't retire before 2004, he would include this new information in his retirement decision, possibly even postpone his retirement. Therefore, one could argue, including the change in our calculations would be the best thing to do. But given he retired prior to when the information became available, it would be incorrect to account for the reform as it would make later retirement more favorable than it actually was when he took the decision to retire early. We have chosen to include some changes occurring between age $a = 0$ and P: changes in the tax- and benefit systems and changes in partner status.

Computing the most accurate future income streams is a compromise between rigid model assumptions and a complex reality. Deciding on the right balance of compromises is an interesting, yet indefinite discussion. Investigating the robustness of the estimation results would be desirable but time demanding. An obvious step further would be to expand to a dynamic model and make individuals reconsider their retirement decision as they age, approaching the dynamic programming framework. Such a development would lead to a significant increase in the complexity of an already complex model. Simplification of the retirement decision enabled us to introduce more complexity in other aspects of the model.

7.6 Computational deliberations

The estimation program is written in **R**, calling **C++** from the R add-on package **Rcpp** which facilitates C++ programming in R. The benefit of programming in **C++** is that it is significantly faster than R. Converting the estimation program from R to C++ reduced the processing time more than 50 times. All estimation, convergence of $p(k)$ and likelihood computation, is done within C++, while the search for optimizing values of α, ϕ and θ is done through the build-in optimization function NLM in R. In order to ensure a reasonable scale of utility we measure consumption in 10,000 DKK, why all the computed income streams and initial wealth accumulations are divided by this amount.

We estimate k on an grid ranging from 0.1 to 3.9 with interval length 0.2. When we solve the consumption problem with EGM we define the wealth grid to contain the 0.2-interval percentiles of a combination of the observed salaries and initial wealth accumulations. As a result, the interval length of the wealth grid varies with large intervals for the periphery wealth accumulations and small intervals in the center of the defined grid. More grid points involve more computations. The higher density of grid points in the intermediate values of wealth optimizes the EGM algorithm as the accuracy of the majority of consumption paths are improved. Whenever a combination of parameters result in a lower log likelihood value, the corresponding estimation of $p(k)$ is chosen as the initial guess of $p(k)$ in the following log likelihood evaluations in order to shorten convergence time.

In order to optimize the converge process of the k-distribution as described in Section 7.1, we apply the method of Successive Over-Relaxation (SOR) which was originally developed in a Gauss-Seidel context of solving linear equation systems. An additional step in the iterative process written in equation 28 is added such that:

$$\hat{p}_{s+1}(k) = \omega \hat{p}_s(k) + (1-\omega)\hat{p}_{s-1}(k)$$

The SOR method exaggerates the change in each iteration whenever $\omega > 1$. The method was tested on simulated data which revealed an optimal value of $\omega = 1.987$, such that the iteration changes are almost doubled. In simulated data, this optimized the algorithm with approximately 24%.

Despite the many efforts to speed up the estimation process, it is still very slow. Depending on the initial assumption about the k-distribution, the chosen convergence criteria and naturally also the number of individuals, each combination of α, ϕ and θ requires up to several hours until the corresponding k distribution and log likelihood value are computed. Finding new methods to speed up the estimation process would be incredible useful and is, indeed, needed. The estimation results should be interpreted with this condition in mind. Due to the limited time available, not all combinations of grids, intervals, parameter values and convergence criteria were tested, why the displayed results aren't optimal with certainty.

8 Estimation Results

The homogeneous parameters α, ϕ and θ are estimated together with the distribution of the heterogeneous variable k. First the model is estimated separately for the two genders, based on a 10% sample of the data set. Then the data is split in both gender and education level and the model is estimated separately for each group. The gender- and education specific estimations constitute the main results of the study why these will be discussed more thoroughly.

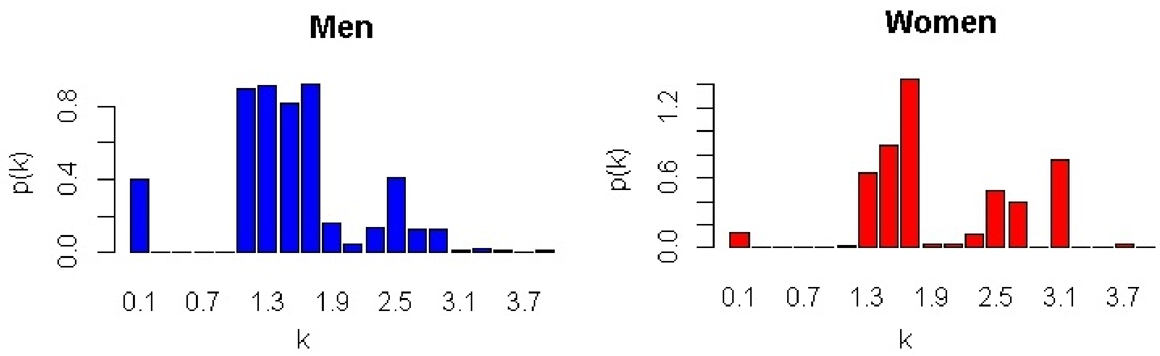
8.1 Gender specific estimation

The homogeneous parameter estimates are presented in Table 6 and the estimated histograms of $p(k)$ are plotted in Figure 14.

Table 6: Homogeneous Parameters - Gender Divided Estimation

	α	ϕ	θ
Men	0.0088	45.32	0.1659
Women	0.0068	39.71	0.1432

Figure 14: Estimated distributions of k



The estimated population distributions of k clearly suggest a significant variation in individuals' preferences for leisure. Both men and women are characterized by k-values above one, suggesting a utility gain in retirement.

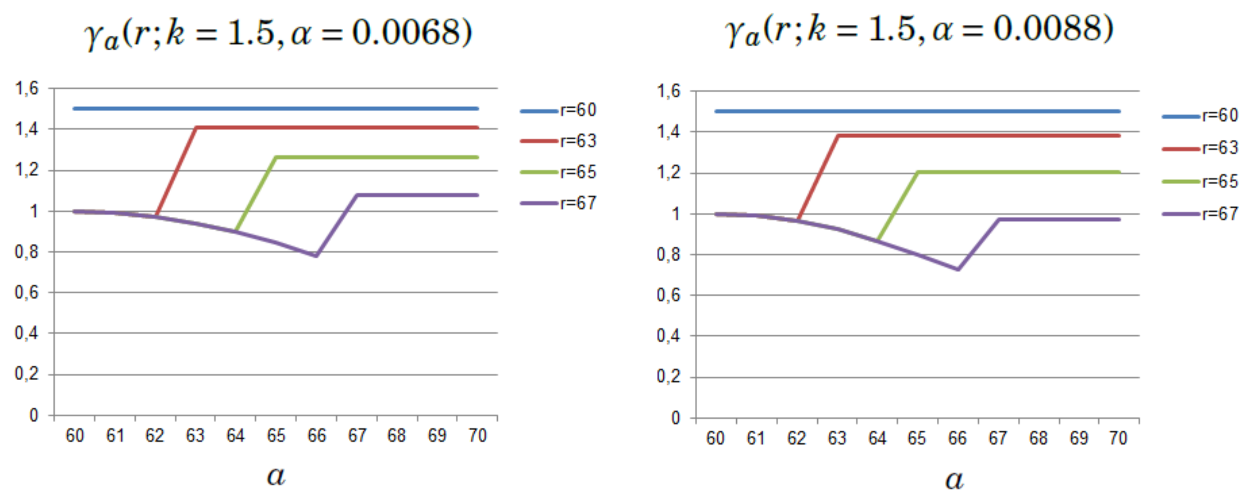
While a large share of both men and women has k values of approximately 1.5, a smaller group of individuals has significantly higher values. A group consisting of approximately 15% of all men and 20% of all women clusters around a k-value of 2.5

while 15% of all women has k-values of 3.

The histogram spike at $k = 0.1$ is positive for both genders but significantly large for men - given the 0.2 interval width it suggests that 8% of all men have k-values close to 0. But we should be careful to make this interpretation, as the estimated spike might be a result of the model specification that does not identify k-values for individuals who retire at the latest (or earliest) possible retirement age, see section 7.1.2.

The gender-specific estimation results suggest that men have higher values of α, suggesting that men to a larger degree than women are pushed into retirement. Understanding what an α estimate of 0.0088 or 0.0068 actually means, Figure 15 show the γ-function for different retirement ages evaluated in these values of α when $k = 1.5$.

Figure 15: γ evaluated for different retirement ages in the estimated α parameters with $k = 1.5$



How much the push-parameter effects the utility of consumption after retirement depends very much on the chosen retirement age. If $\alpha = 0.0088$, the utility level of consumption is decreased with one third if the retirement age is 67 compared to 60. So the estimated values of α has a quite large effect on the retirement decision and the effect is largest for men.

The estimated values of θ, the subjective time preference rate are in line with previous studies that estimate the magnitude of θ to be 0.1-0.2, see Andersen et al. (2008).

8.2 Gender- and education specific estimation

The sample is divided in 10 gender- and education specific groups which are estimated separately. Education levels are split in unskilled, vocational, short-, medium- and long tertiary. Approximately 1,5% of the sample were registered with unknown education level and this group was excluded from the analysis.
The estimated heterogeneous parameter k affects the retirement behavior *together* with the homogeneous parameters why they should all be interpreted together. The homogeneous parameter estimates are listed in Table 7 while the estimated distributions/histograms of $p(k)$ are plotted in Figure 16.

Table 7: Estimation results of homogeneous parameters

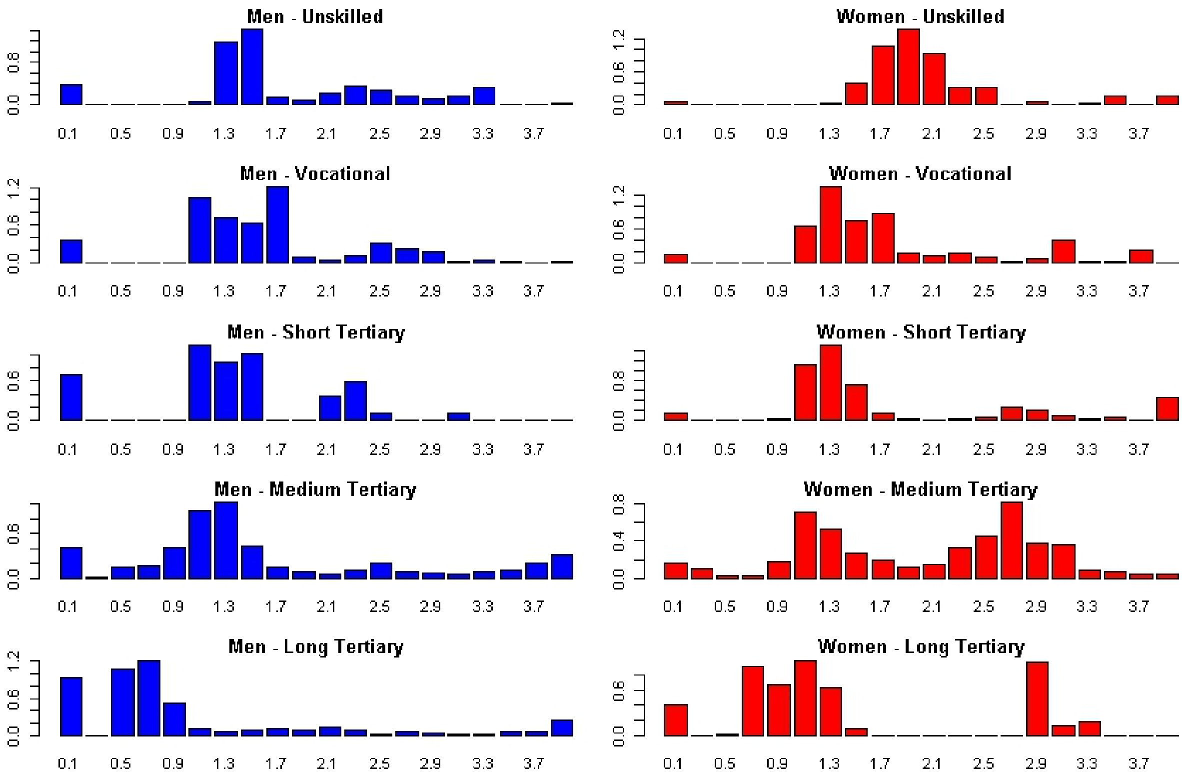
	Men			**Women**		
	α	ϕ	θ	α	ϕ	θ
Unskilled	0.0065	40.21	0.1178	0.0046	44.41	0.1661
Vocational	0.0091	43.22	0.1497	0.0084	49.76	0.1647
Short Tertiary	0.0104	46.00	0.2440	0.0070	38.77	0.1307
Medium Tertiary	0.0088	27.92	0.1031	0.0107	27.85	0.1371
Long Tertiary	0.0101	26.23	0.0765	0.0068	20.98	0.0441

The bi-modular distribution of preferences for leisure is extremely significant for women with longer education, while the bi-modular distribution is more pronounced for men with shorter education compared to men with long education.

The estimated values of θ, denoting the subjective time preference rate, are in line with previous studies, see Andersen et al. (2008), that estimate the magnitude of θ to be 0.1- 0.2. Individuals with less education generally seem to have higher values of θ, implying that these are more impatient. The estimated levels of θ are much more reasonable compared to the estimates in Arnberg and Stephensen (2015) where the estimated θ values varying from 0.2-0.5. This significant change in parameter estimates of θ is presumably a result of the difference in credit market assumptions. Arnberg and Stephensen (2015) assume identical age-dependent interest rates for debt and deposits while the current model assumes asymmetry such that the interest rate on debt equals the age-dependent interest specification in Arnberg and Stephensen (2015) while the interest rate on deposit is assumed constant and equal to 4.75%. This assumption makes it less favorably to accumulate wealth in old age, making dis-saving behavior more probable for lover rates of θ.

The parameter ϕ is lowest for the higher educations, indicating that the model fits the groups with less education better as the estimated model is less stochastic compared to the higher educations. Another interpretation is that individuals with higher

Figure 16: Estimated distributions of k



education are less likely to follow financial incentives, why the unobserved heterogeneity, ϵ_r has more influence on their retirement decision compared to individuals with lower educations. However, this result seem contra intuitive. The sample sizes differ significantly for the different groups, see Table 4. The highest ϕ estimates tend to apply to the largest groups, indicating that the retirement decision is more precisely modeled in the large sample groups. The smallest groups, however, consists of more than 1000 individuals.

The push-parameter α is consistently higher for men that women but not consistently increasing/decreasing with education level. Surprisingly, the estimation results suggest that unskilled individuals are the ones with the smallest α estimates, suggesting that unskilled individuals are less exposed to attrition which we know is untrue. Now, the k distribution for the unskilled is also characterized by high levels of k. A high k reflects high preferences for leisure, or put differently, dislike of working. As the push and pull factors act concomitantly, the effect of high k-values on the retirement decision are outweighed by a corresponding effects of low α's and the other way around. If the model was estimated with $\alpha = 0$, the estimated k distributions would be significantly higher. Never the less, the estimation results suggest that the retirement decision is more voluntary for unskilled individuals than for individuals with e.g. higher education.

The estimated k-distribution for women with medium tertiary education is especially interesting as it suggest an equal division of the group in two halves: one half with k-values of approximately 1.2 and one half with k-parameters of around 3. Women with medium tertiary education are typically employed as e.g. nurses and teachers. The estimated α parameter for this group is higher than for the remaining groups with $\alpha = 0.0107$, indicating a high level of attrition which could be both physical and mental. The high level of α should again be interpreted in the context of a highly bi-modular distribution of k. Those with a combination of a low k and high α are forced into retirement despite the fact that they almost don't appreciate being retired. Those with a combination of a high k and high α face a more voluntary retirement decision where an early retirement decision is rewarded both in terms of a significant jump in utility and in terms of avoiding a potential high degree of attrition.

Women with long tertiary education are extremely divided in their preferences for leisure. The majority have k-values vary from 0.7-1.3 but then approximately 20% of the group are estimated to have $k = 3$, suggesting a strong dislike of working or a strong appreciation of free time compared to the remaining 80% of the group. The corresponding push parameter is low compared to the other groups with $\alpha = 0.0068$ suggesting a relatively low level of attrition.

To men with long tertiary education, the retirement decision is a choice between the devil and the deep blue sea. k values below 1 (corresponding to a utility loss when retiring) and a high value of α suggests that they are forced into retirement: they don't want to retire but as their utility is permanently worsened the longer they stay in the labor force, they are forced to retire eventually. A high level of α might not only reflect physical attrition but might also reflect worsening mental health condition such as stress.

In order to ease comparability, all the estimated distributions of k are also plotted together in one graph for each gender, men in Figure 17 and women in Figure 17. Plotting the estimated distributions of all education group together emphasizes how the k distributions shifts to the right for higher educational groups, both for men and women but most pronounced for women.

If an individual has high preferences for leisure it might indicate that he simply values his free time a lot - maybe because he likes to spend time with his wife, grand

Figure 17: Estimated distributions of k - all male education groups

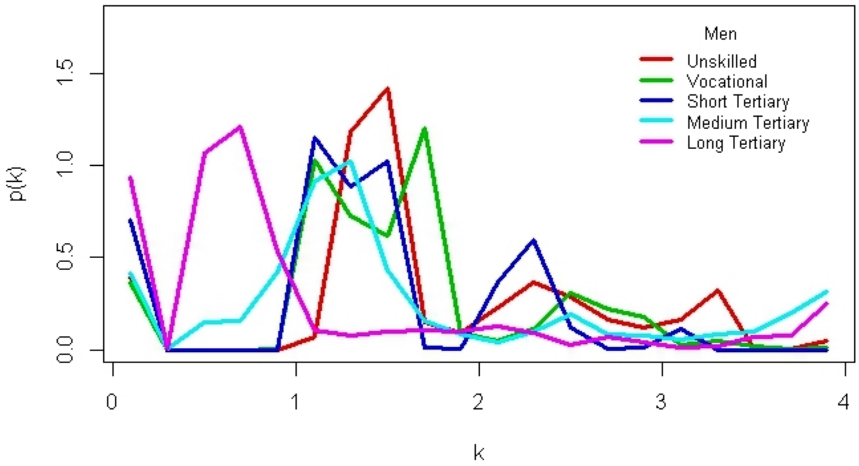
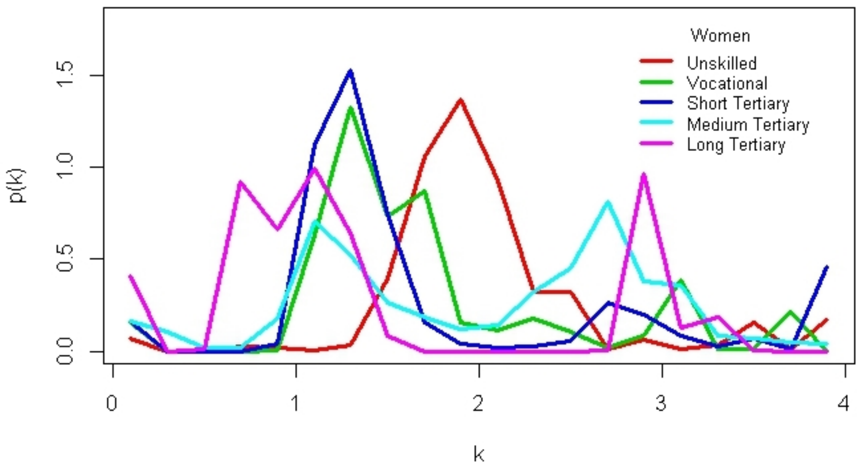


Figure 18: Estimated distributions of k - all female education groups



children or a passionate hobby. But it might also suggest that he just dislikes working, making the relative value of free time compared to work more valuable. The fact that the level of k decreases with longer education might therefore also suggest that individuals with low education dislike working more than individuals with long education. The only groups with k-distributions considerably below 1 are men and women with long tertiary educations. It makes good sense that individuals with high education value their job more, maybe because their self-identity is closely tied to work or because they simply enjoy working.

In contrast to women with medium- and long tertiary education, uskilled women have very little evidence of bi-modularity and the estimated distribution is almost uniform. The group tend to have large k-parameters compared to the remaining group, varying from 1.5 to 2.5.

8.3 In-sample forecast

Given our estimation results, what can we say about the model fit? We can estimate the expected distribution of retirement ages given our model estimates and data and compare these with the actual distribution of retirement ages. We do a so-called in-sample forecast. In order to calculate the expected distribution of retirement ages we must compute, for each individual, his probability to retire at any of the ages 60,..,67. The probability that individual j retires at age r is:

$$\begin{aligned} Pr(r|x^j) &= \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) P^j(k|r^j;\alpha,\phi,\theta)dk \\ &= \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) \frac{\pi\left(r^j|k,\alpha,\phi,\theta,x^j\right)p(k)}{P(r^j;\alpha,\phi,\theta)}dk \end{aligned} \tag{37}$$

Summing op the total probability of retiring at a given age for all individuals give us the expected number of retirements the given year. Let N_r denote the expected number of individuals who retire at age r:

$$n_r = \sum_{j=1}^{n} Pr(r|x^j)$$

Table 8 states the actual and predicted retirement age distributions for men while Table 9 state those for women:

Table 8: Actual and predicted retirement age distributions for men

	All		Unskilled		Vocational		Short Tert.		Medium Tert.		Long Tert.	
r	Act	Pred	Act	Pred	Act	Pred	Act	Pred	Act	Pred	Act	Pred
60	**18.9**	18.1	**22.4**	22.4	**20.4**	19.4	**15.4**	14.2	**13.6**	12.4	**7.6**	6.2
61	**8.5**	12.8	**9.9**	11.8	**9.1**	14.9	**7.0**	12.4	**6.6**	11.8	**3.2**	6.8
62	**18.2**	13.2	**17.8**	14.8	**20.9**	13.6	**15.1**	11.8	**16.5**	11.8	**9.2**	7.5
63	**10.4**	11.5	**9.9**	10.8	**10.9**	12.0	**11.3**	11.2	**10.9**	12.2	**8.6**	9.8
64	**7.3**	10.6	**6.2**	9.4	**7.2**	10.5	**8.0**	10.5	**8.9**	12.3	**9.5**	12.5
65	**12.9**	8.3	**10.3**	6.9	**11.5**	7.7	**13.8**	8.8	**16.8**	10.7	**24.5**	13.5
66	**5.9**	8.8	**4.9**	7.4	**5.2**	7.7	**5.6**	9.6	**8.0**	11.2	**9.9**	16.6
67	**17.8**	16.6	**18.6**	16.5	**14.9**	14.2	**23.8**	21.6	**18.6**	17.6	**27.6**	27.1

The in-sample forecast seem to average out the kinks in actual retirement ages. This is clearly illustrated in Figure 19 where the actual and predicted retirement ages are plotted for the weighted averages of the two genders.

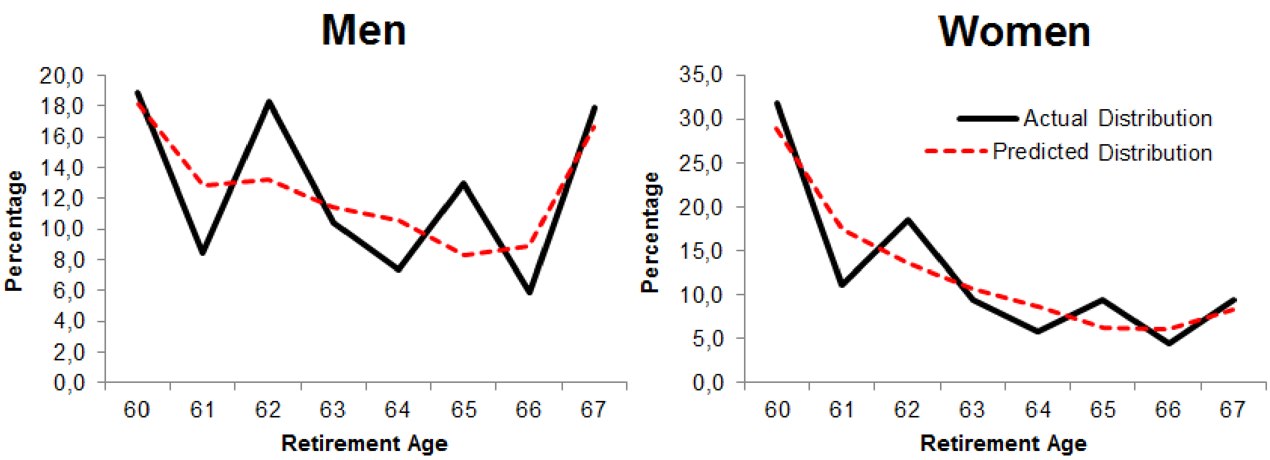
The kinks in the actual retirement ages reflect the financial incentives to retire at age 60, 62, 65 and 67. Whereas the predictions are very accurate for age 60 and 67, they fail at capturing the kinks at age 62 and 65. The result is an underestimated

Table 9: Actual and predicted retirement age distributions for women

	All		Unskilled		Vocational		Short Tert.		Medium Tert.		Long Tert.	
r	Act	Pred	Act	Pred	Act	Pred	Act	Pred	Act	Pred	Act	Pred
60	**31.8**	29.0	**38.5**	35.0	**30.8**	28.8	**22.2**	20.4	**27.2**	23.5	**11.7**	9.8
61	**11.2**	17.4	**12.0**	18.3	**11.4**	17.2	**7.8**	14.6	**11.2**	18.3	**3.9**	9.9
62	**18.5**	13.7	**17.0**	12.9	**19.6**	14.1	**18.4**	13.0	**20.0**	15.0	**14.6**	10.4
63	**9.4**	10.7	**7.8**	9.3	**9.7**	11.1	**12.5**	11.7	**11.1**	12.3	**9.7**	11.7
64	**5.8**	8.7	**4.8**	7.2	**6.0**	8.9	**7.2**	10.9	**5.7**	9.9	**10.8**	13.0
65	**9.5**	6.2	**8.0**	5.2	**9.1**	5.8	**12.3**	8.5	**10.4**	7.1	**20.5**	12.4
66	**4.4**	6.1	**3.9**	5.3	**4.0**	5.8	**5.7**	9.2	**5.4**	6.2	**7.9**	13.3
67	**9.4**	8.2	**7.9**	6.8	**9.4**	8.3	**13.9**	11.8	**9.1**	7.8	**20.8**	19.5

Note: The total predicted distribution for men and women are computed as the weighted averages of the education specific predictions.

Figure 19: Predicted vs. actual retirement distributions



number of retirements at age 62 and 65 and a corresponding overestimation of retirements at age 62 and 65. This could indicate that the computed income streams underestimate the financial incentives that applies to the retirement ages 62 and 65.

The fact that the predicted retirement age distribution is too smooth is also related to the small estimates of ϕ. The predicted retirement ages are more accurate for those with the highest ϕ estimates. A high value of ϕ indicates that the unobserved heterogeneity becomes less influential on the retirement decision and the financial incentives become more important. A low value of ϕ, on the other hand, results in a decision that is more random and more affected by non-financial incentives.

Notice that equation 37 applies the individual r^j-conditional probability to compute the in-sample forecast. Thereby , information of each individuals' actual retirement behavior is applied when forecasting their predicted retirement behavior. Each person's individual k-distribution is estimated from the knowledge of his observed retirement age, and this individual-specific information is useful in the forecast. However, for out-of sample forecast that concerns another group of individuals, we're not able to use this individual-specific information why the individual distributions of k are replaced with the population distribution such that

$$Pr(r|x^j) = \int_{\mathcal{K}} \pi\left(r|k,\alpha,\phi,\theta,x^j\right) p(k)dk \tag{38}$$

This loss of information obviously means that out-of-sample forecast are more imprecise compared to in-sample forecasts. However, as the estimated value of $p(k)$ is in fact just the average of all person's individual k-distributions, the method described in equation 37 and 38 produce identical retirement distribution for in-sample forecasts.

8.4 Further research proposals

The estimated distributions suggest that there is a significant variance in individuals' preferences for leisure in the context of retirement. Such a significant finding deserves, in my humble and strongly biased opinion, further investigations which were omitted in this analysis due to time limitations.

During the initial estimation phase I tried to estimate with different combinations of k-intervals and step sizes. I chose the interval $0.1-3.9$ with stepsize 0.2 as a compromise between accuracy and computation time. However, changing the underlying k-grid resulted in significant changes of parameters estimates and estimated distribution. In order to test the sensitivity of the results with respect to the selected grid-intervals it would be interesting to do a cross-validation test on the estimation result with different k-intervals.

Many of the estimated histograms have significant spike at $k = 0.1$, especially for men. These might very well be explained by the identification issue described section 7.1.2 caused by individuals who retire at the earliest possible retirement age. Expanding the model into including retirement at age 59 might help on this identification issues. $k = 0$ would then imply retirement at age 59 which almost no one chooses, why the spikes at the left distribution tails would probably disappear.

The relatively simple histogram estimation technique applied in this study could relatively easily be improved. There exist an extensive literature on non-parametric estimation techniques, e.g. kernel estimation methods, which would provide more accurate heterogeneity estimations.

From the maximum likelihood estimation we are able to compute statistical inference such as standard errors which could easily be computed numerically, e.g. with the package **numDeriv** in R. However, as the iterative convergence of k is nested within the maximum likelihood evaluation, the computation of the Hessian is both time demanding and it dependent on the choice of k-grid, convergence criteria, EGM wealth criteria etc. The computed standard errors would, however, be a useful indicator of the accuracy of the estimation parameters, but were not included in the analysis due to time limitations.

The model relies on several critical assumptions which were outlined in Section 7.5 together with suggested solutions to avoid them. An issue that should be thoroughly investigated is the correlation between the heterogeneous parameter k and the background variables x_j. Allowing for interdependence between the heterogeneity and x_j would be an important contribution to the existing model.

As described in the literature study in section 3.2, previous studies find evidence of important effects of health and joint retirement on the retirement decision. With the current push/pull model framework the effects of e.g. health and joint retirement are hard to identify. see section 3.3.1. An obvious development of the model would therefore be to account for these effects and investigate how they affect the measured heterogeneity in leisure preferences. Another obvious development would be to introduce job uncertainty. As the current model already applies an EGM solution, the introduction of dynamic programming seems straight forward.

Whereas the results alone reveal interesting information about heterogeneity in retirement behavior, the estimated model also constitutes a useful tool for policy analysis which has not been utilized at all in this study. An obvious next step would be to apply the model on policy experiments and out-of-sample forecasts and investigate how accounting for heterogeneity affects these analyses.

9 Conclusion

The main purpose of this study is to estimate heterogeneity in the retirement decision with a proposed non-parametric estimation technique. The estimated results show significant variation in people's preferences for leisure and suggest a bi-modular distribution with a large group with relatively low preferences for leisure and a small group of individuals with significant larger preferences for leisure.

The model is estimated separately for ten gender- and education specific groups and we find significant variation in the estimation results for the different groups. The bi-modular distribution of preferences for leisure is extremely significant for women with longer education, while the bi-modular distribution is more pronounced for men with shorter education compared to men with longer education. With women having higher values of the pull-parameter, k and lower values of the push-parameter α, the estimation results suggest that women's retirement decision is more voluntary compared to men. Individuals with lower educations tend to have higher preferences for leisure and are less patient, reflected in lower rates of time preferences, compared to individuals with long education. The estimation results also suggest that - due to low ϕ estimates - the unobserved heterogeneity has a large impact on the retirement decision resulting in imprecise in-sample forecasts.

The estimated model assumes different interest rates such that the interest rate on deposits is constant while the interest rate on debt increases with the debt takers age. How this affects the estimation results is hard to say, but we do find significantly lower and more credible estimates of the subjective rates of time preferences compared to previous comparable studies with more simple (and unrealistic) credit market assumptions.

The estimated heterogeneity and the tendency of bi-modular distributions is a highly interesting discovery that contributes to a deeper understanding of peoples retirement behavior. The model has a great potential for further developments that could contribute with a further understanding of heterogeneity in the retirement decision.

10 Appendix

10.1 Derivation of the Logit Model

Consider individual j. As we don't know the error terms $\epsilon^j = (\epsilon_1^j, ..., \epsilon_P^j)$ we treat them as random.

The probability that individual j chooses retirement age r is given by

$$\begin{aligned} Pr(r) &= Pr(U_0(r)^j > U_0(a)^j \ \forall a \neq r) \\ &= Pr(\phi V_0(r)^j + \epsilon_r^j > \phi V_0(a)^j + \epsilon_a^j \ \forall i \neq a) \\ &= Pr(\epsilon_r^j - \epsilon_a^j < \phi V_0(r)^j - \phi V_0(a)^j) \ \forall a \neq r) \end{aligned} \tag{39}$$

This probability is a cumulative distribution, i.e. the probability that each random term $\epsilon_r^j - \epsilon_a^j$ is below the observed quantity $\phi V_0(r)^j - \phi V_0(a)^j$. In order to measure this probability we need do some assumptions about the distribution of the error term, $f(\epsilon)$. Using the density $f(\epsilon)$, the cumulative probability in equation 39 can be written as

$$\begin{aligned} Pr(r) &= Pr(\epsilon_r^j - \epsilon_a^j < \phi V_0(r)^j - \phi V_0(a)^j \ \forall a \neq r) \\ &= \int_\epsilon I\left(\epsilon_r^j - \epsilon_a^j < \phi V_0(r)^j - \phi V_0(a)^j\right) f(\epsilon^j) d\epsilon^j \end{aligned} \tag{40}$$

where $I(\cdot)$ is the indicator function equal to 1 when the expression in parentheses is true and 0 otherwise. As you can see, this is a multidimensional integral over the density of the unobserved portion of utility, $f(\epsilon^j)$. As ($U_0(r)$ is unobserved, measuring the distribution of the error terms is impossible why we must assume some distribution. Different assumed distributions leads to different discrete choice models. If we assume that ϵ follow a normal distribution, we will get a Probit model. If we assume that ϵ follow an extreme value distribution, we will arrive at the Logit model. Logit and Probit models are the most prevalent models in similar studies, and the two specifications tend to deliver very similar results. This is because the normal distribution resembles the extreme value distribution except from the extreme value distribution having slightly fatter tails. However, assuming that ϵ is iid. extreme value distributed, the multidimensional integral in 40 has a nice, closed form solution, why this model is less computationally demanding. The extreme value distribution has variance $\frac{\pi^2}{6}$ why we are implicitly normalizing the scale of utility. We multiply $V_0(r)^j$ by ϕ in order to fit it on the normalized utility scale. The mean of the distribution is not zero, but this doesn't matter since the utility scale is ordinary: we don't care about the level of utility itself, only the differences in utilities.

The joint density of ϵ^j is denoted $f(\epsilon^j)$. With this density, we are able to make probabilistic statements about individual j deciding to retire at age a:

$$
\begin{aligned}
Pr^j(r) &= Pr(U_0(r)^j > U_0(a)^j \ \forall a \neq r) \\
&= Pr(\phi V_0(r)^j + \epsilon_r^j > \phi V_0(a)^j + \epsilon_a^j \ \forall i \neq a) \\
&= Pr(\epsilon_r^j < \epsilon_a^j + \phi V_0(r)^j - \phi V_0(a)^j \ \forall a \neq r)
\end{aligned} \tag{41}
$$

Suppose we knew the value of the error terms ϵ^j, then we would know the choice probabilities $Pr^j(r)$ conditional on the information about ϵ^j, $Pr^j(r)|\epsilon^j$. Define $B_a = \epsilon_a^j + \phi V_0(r)^j - \phi V_0(a)^j$, this gives us:

$$
\begin{aligned}
Pr^j(r) \mid \epsilon^j &= Pr(\epsilon_1^j < B_1 \& \epsilon_2^j < B_2 \& ...) \\
&= Pr(\epsilon_a^j < B_a) \forall a \neq r
\end{aligned} \tag{42}
$$

Assume that ϵ^j are iid extreme value distributed. The extreme value distribution is given by the density function

$$
f(\epsilon^j) = e^{-\epsilon^j} e^{-e^{-\epsilon^j}}
$$

And cumulative distribution

$$
F(\epsilon^j) = e^{-e^{-\epsilon^j}}
$$

Due to the independence of the error terms, the probability in 42 is just the product of the individual densities:

$$
\begin{aligned}
Pr^j(r) \mid \epsilon^j &= Pr(\epsilon_1^j < B_1) \cdot Pr(\epsilon_2^j < B_2) \cdot \ldots \cdot Pr(\epsilon_P^j < B_P) \\
&= e^{-e^{-B_1}} \cdot e^{-e^{-B_2}} \cdot \ldots \cdot e^{-e^{-B_P}} \\
&= \prod_{a \neq r} e^{-e^{-B_a}}
\end{aligned}
$$

Obviously ϵ^j is not known, and hence we must arrive at the unconditional probability by integrating ϵ^j out - we evaluate the integral of $Pr(a_r = a|\epsilon^j)$ over all possible values of ϵ^j weighted by its density:

$$
\begin{aligned}
Pr(r) &= \int_{-\infty}^{\infty} Pr^j(r) \mid \epsilon^j f(\epsilon^j) d\epsilon^j \\
&= \int_{-\infty}^{\infty} \left(\prod_{a \neq r} e^{-e^{-B_a}} \right) e^{-\epsilon^j} e^{-e^{-\epsilon^j}} d\epsilon^j
\end{aligned}
$$

We wish to evaluate the above integral, and the first step is to remove the restriction $a \neq r$ by multiplying by $e^{e^{-\epsilon^j}} \cdot e^{e^{\epsilon^j}} = 1$:

$$Pr(r) = \int_{-\infty}^{\infty} \left(\prod_{a=1}^{P} e^{-e^{-B_a}} \right) e^{-\epsilon^j} e^{-e^{-\epsilon^j}} e^{e^{\epsilon^j}} d\epsilon^j$$
$$= \int_{-\infty}^{\infty} \left(\prod_{a=1}^{P} e^{-e^{-B_a}} \right) e^{-e^{-\epsilon^j}} d\epsilon^j$$

Now consider the product. It can be written as

$$\prod_{a=1}^{P} e^{-e^{-B_a}} = e^{-\sum_{a=1}^{P} e^{-B_a}}$$
$$= e^{-\sum_{a=1}^{P} e^{-(\epsilon_a^j + \phi V_0(r)^j - \phi V_0(a)^j)}}$$
$$= e^{-e^{-\epsilon_a^j} \sum_{a=1}^{P} e^{-(\phi V_0(r)^j - \phi V_0(a)^j)}}$$

Define $Q = \sum_{a=1}^{P} e^{-(\phi V_0(r)^j - \phi V_0(a)^j)}$, this gives us

$$Pr(r) = \int_{-\infty}^{\infty} e^{-e^{-\epsilon_a^j} Q} e^{-e^{-\epsilon^j}} d\epsilon^j$$

where Q is independent of the variable of integration ϵ^j. Now we do integration by substitution with the change of variable $y = e^{-\epsilon^j}$. This transformation maps $[-\infty, \infty]$ onto $[0, \infty]$ with the inverse transformation being $\epsilon^j = -\ln y$. The Jacobian of the inverse transformation is $J = \frac{d\epsilon^j}{dy} = \frac{-1}{y}$. Since $y > 0$ the absolute value of the Jacobian is $|J| = \frac{1}{y}$. Under the change of variable we reach

$$Pr(r) = \int_{0}^{\infty} e^{-Qy} \cdot y \cdot |J| dy$$
$$= \int_{0}^{\infty} e^{-Qy} dy$$
$$= \left[-\frac{1}{Q} e^{-Qy} \right]_{0}^{\infty}$$
$$= \frac{1}{Q}$$
$$= \frac{1}{\sum_{a=1}^{P} e^{-(\phi V_0(r)^j - \phi V_0(a)^j)}}$$
$$= \frac{1}{e^{-\phi V_0(r)^j} \sum_{a=1}^{P} e^{\phi V_0(a)^j}}$$
$$= \frac{e^{V_0(r)^j}}{\sum_{a=1}^{P} e^{V_0(a)^j}}$$

10.2 Banach's Fixed Point Theorem

Definition 10.2.1. *Let* (X, d) *be a metric space. Then a map* $T : X \to X$ *is called a contraction mapping on* X *if there exists* $q \in [0,1)$ *such that* $d(T(x), T(y)) \leq q d(x, y)$ *for all* x*,* $y \in X$*,* $y \neq x$*.*

Theorem 10.2.1 (Banach's Fixed Point Theorem)**.** *Let* (X, d) *be a non-empty complete metric space with a contraction mapping* $T : X \to X$*. Then* T *admits a unique fixed-point* $x*$ *in* X *(i.e.,* $T(x^*) = x^*$*). Furthermore,* x^* *can be found as follows: start with an arbitrary element* x_0 *in* X *and define a sequence* x_s *by* $x_s = T(x_{s-1})$*, then* $x_s \to x^*$*.*

Bibliography

Allenby, Greg M. and Peter E. Rossi (1999), “Marketing models of consumer heterogeneity.” *Journal of Econometrics 89, pp. 57-78*.

An, Mark, Bent Christensen, and Nabanita Gupta (1999), “A bivariate duration model of the joint retirement decisions of married couples.” *CLS Working Papers 99–10 - Centre for Labour Market andSocial Research, Denmark*.

An, Mark, Bent Christensen, and Nabanita Gupta (2004), “Multivariate mixed proportional hazard modelling of the joint retirement of married couples.” *Journal of Applied Econometrics, vol 19 issue 6*.

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and Elisabet Ruström (2008), “Eliciting risk and time preferences.” *Econometrica 78(3) pp. 583-619*.

Arnberg, Søren and Peter Stephensen (2013), “Pension og tilbagetrækning - ikke-parametrisk estimation af heterogenitet.” *DREAM working paper 2013:2*.

Arnberg, Søren and Peter Stephensen (2015), “økonomiske incitamenter, nedslidning og tilbagetrækning - semi-parametrisk estimation af heterogenitet og aldersbetingede ønsker om tilbagetrækning.” *DREAM working paper*.

Azar, Samih (2010), “Bounds to the coefficient of relative risk aversion.” *Banking and Finance Letters Volume 2, Issue 4, pp. 391-398*.

Belloni, Michele (2008), “The option value model in the retirement literature: The trade-off between computational complexity and predictive validity.” *ENEPRI Research Report No. 50*.

Bingley, Paul, Nabanita Gupta, and Peder Pedersen (2004), “The impact of incentives on retirement in denmark.” *Review of Economic Studies (2005) 723, pp.395-427*.

Bingley, Paul and Gauthier Lanot (2007), “Public pension programmes and the retirement of married couples in denmark.” *Journal of Public Economics 91 (2007), pp. 1878–1901*.

Blau, David (1998), “Labor force dynamics of older married couples.” *Journal of Labor Economics vol. 16, no.3*.

Boskin, Michael (1977), “Social security and retirement decision.” *Economic Inquiry 15, pp. 1-25*.

Carroll, Christopher (2006), “The method of endogenous gridpoints for solving dynamic stochastic optimization problems.” *Economic Letters, 93(3), pp. 312-320*.

Danø, Anne, Mette Ejrnæs, and Leif Husted (2004), "Gender differences in retirement behaviour." *AKF memo*.

Feldman, Daniel (1994), "The decision to retire early: A review and conceptualization." *The Academy of Management Review Voume. 19. No. 2, pp. 285-311.*

Finansministeriet (2014), *Budgetoversigt 3, 2014*. Finansministeriet.

French, Eric (2004), "The effects of health, wealth, and wages on labour supply and retirement behaviour." *Review of Economic Studies (2005) 723, pp.395-427*.

Gustman, Alan and Steinmeier Thomas (1986), "Integrating retirement models." *Econometrica 3(54), pp. 555-584*.

Gustman, Alan and Steinmeier Thomas (2009), "Integrating retirement models." *NBER Working Papers 15607, National Bureau of Economic Research, Inc.*

Hausman, Jerry and Daniel McFadden (1984), "Specification tests for the multinomial logit model." *Econometrica, Vol. 52, Issue 5 pp.1219.1240*.

Hausman, Jerry and David Wise (1985), "Social security, health status, and retirement." In *Pensions, Labor, and Individual Choice* (David Wise, ed.), University of Chicago Press, pp. 159-191, Chicago.

Honoré, Bo and Aureo De Paula (2013), "Interdependent durations in joint retirement." *Working Paper CWP05/13, The Institute for Fiscal Studies Department of Economics, UCL (CEMMAP).*

Jørgensen, Thomas (2014), *Leisure Complementarities in Retirement*. Ph.D. thesis, University of Copenhagen.

Kramp, Paul, Jane Lohff, and Jens Maltbæk (2012), *Monetary Review 2012 1st Quarter Part 1 - Pension Savings*. Danmarks Nationalbank.

Larsen, Mona and Nabanita Gupta (2010), "A structural retirement model." *Health Economics Volume 19(7), pp. 792-813*.

McGarry, Kathleen (2004), "Health and retiremen: Do changes in healt affect retirement expectations?" *Journal of Human Resources. 39, (3), pp. 624-648*.

Mitchell, Olivia. and Gary Fields (1984), "The economics of retirement behavior." *Journal of Labor Economics 2(1), pp. 84-105.*

Robin Lubsdain, James Stock and David Wise (1992), *Three Models of Retirement: Computational Complexity versus Predictive Validity - in D. Wise (ed.) TTopics of Economics of Aging*. University of Chicago Press, pp. 21-60.

Rust, John (1987), “Optimal replacement of gmc bus engines: An empirical model ofharold zurcher.” *Econometrica, 55, pp. 999–1033*.

Rust, John (1989), *A Dynamic Programming Model for Retirement Behavior - in D. Wise (ed.) The Economics of Aging*. University of Chicago Press, pp. 359-398.

Stephensen, Peter (2009), “Dream’s livsforløbsmodel - model og algoritme.” *DREAM*.

Stock, James and David Wise (1990), “Pensions, the option value of work and retirement.” *Econometrica, Vol. 58, Issue 5 pp.1151-1180*.

Train, Kenneth (2007), “A recursive estimator for random coefficient models.” *Working Paper, University of California*.

Train, Kenneth (2009), *Discrete Choise Methods With Simulation*. Cambridge University Press.