Retirement Behavior of the Swedish Notch Babies: Evidence from the Job Episodes in the Survey of Health, Ageing and Retirement in Europe

Haodong Qi*

Abstract

The 1994 Swedish pension reform introduced cohort differentials in benefit accounting. Those born in 1938, the "Swedish Notch Babies", were the first recipients whose benefits were partially computed by the Notional Defined Contribution scheme, while older cohorts remained unaffected. This paper examines the aftermath of the reform by analyzing the differences in retirement behavior between the 1937 cohort and the Notch Babies. Both static and dynamic programming retirement models are implemented using Hierarchical Bayesian Estimation. Retirement propensity is measured by the required rate of replacement (R^*). It reflects the level of pension entitlements relative to labor earnings necessary in order for an individual to retire. Large R^* implies low retirement propensity, and vice versa. The empirical results are based on the working life history in the Swedish Survey of Health, Ageing and Retirement in Europe (SHARELIFE). The estimated hyper-parameter (R^*) in both static and dynamic programming models are nearly identical, 0.76 and 0.73, respectively due to the low discounting factor estimated by the dynamic model, $\beta = 0.31$. At the individual-level, R^* differs considerably across cohorts. For the 1937 cohort, tertiary education has large and significant effect on R^*, while gender and health have no impact, ceteris paribus. However, among the Notch Babies, the positive effect of higher education on R^* is reversed, while R^* is much higher for men than women, ceteris paribus. Such cohort differences are identical in both static and dynamic models. The implication of the analysis is three-folded. First, future utility flows have little impact on the retirement decision for both the unaffected and notch cohorts. Secondly, the effect of the reform at the population level is negligible. Finally, the reform encouraged those with completed tertiary education to retire early, while simultaneously prompting male workers to prolong working life.

Keywords: Retirement Behavior, Pension Reform, Dynamic Programming, Hierarchical Bayesian Estimation, Required Rate of Replacement

JEL Classification: H31, H55, J26

*Ph.D. Candidate at Centre for Economic Demography, Lund University, Sweden. Tel: +46 (0)46 22 20 498. Cell: +46 (0)73 72 72 489. Email: Haodong.Qi@ekh.lu.se

Introduction

”NOTCH BABIES” originally refers to the birth cohorts that were adversely affected by the 1977 amendments to the U.S. Social Security Act. These amendments abruptly lowered the prospective retirement benefits for those born 1917-21, while leaving older cohorts unaffected (Krueger and Pischke, 1992). This differential is called a ”notch” and can also be observed in Sweden over the 1994 Pension Reform. The cohorts who were left unaffected are those born in 1937 or earlier. The 1938 cohort was the first recipient whose benefit was computed according to the new rules; therefore, they became the notch generation in the 1994 Swedish Pension Reform, the ”Swedish Notch Babies”.

It has been widely believed that the downward trend of the old-age labor supply during the second half of the 20th century, particularly for men, is attributable to the establishment of generous pension systems in most developed countries. If such a causal link holds, one shall expect that the trend should be reversed if the system becomes less generous. Nonetheless, evidence from the U.S. notch babies show that the labor supply continued to decline even though their social security benefits were substantially lowered relative to the proceeding generations (Krueger and Pischke, 1992).

In Sweden, the mean retirement age exhibited a linear decline over the cohorts born between year 1922 and 1937; however, the downward trend was reversed for the cohorts born between 1938 and 1942. Some argued that such cohort differentials were attributable to number of policy amendments. Lowering the compulsory retirement age from 67 to 65 by year 1976 and launching the early retirement scheme during the 1998s might explain the drop in retirement age across those born between 1922 and 1937. And the increase in the mean exit age for those born after year 1937 might attribute to the 1994 pension reform and the 2003 Employment

Protection Act [1] (Karlsson and Olsson, 2012). Such aggregate patterns do imply that the old age labor supply, at least at the extensive margin, responds to policy and institutional amendments. Furthermore the causal link between pension system, social security and old-age labor supply might be held, which contradicts the evidence from the US notch babies. This motivates the first question in this paper. Did the 1994 reform of the Swedish pension system stimulate the old-age labor participation? Specifically, did retirement behavior differ between the Swedish Notch Babies and those who were unaffected by the reform?

Secondly, to understand the mechanisms of retirement behavior, it is necessary to clarify the effects of policy amendments on income and pension benefits (both current and prospective), as well as on the associated value of either working or retiring. Such crucial components are, however, missing in the aggregate pattern. Hence, this paper stresses on the retirement behavior at the micro level and verifies whether there is conjecture between the micro behavior and the macro emergence.

Thirdly, this paper introduces a new measure of retirement propensity, the required rate of replacement - R^*. This measure reflects the level of pension entitlements relative to labor earnings necessary in order for an individual to retire. A large R^* implies low retirement propensity, and vice versa. Such a measure implies the extent to which people are willing or unwilling to retire. It also sheds some light on why retirement age varies across individuals.

Many previous studies neglect social and demographic determinants of retirement (Karlstrom et al., 2004; Lumsdaine et al., 1990; Stock and Wise, 1990). Nevertheless, besides the fact that retirement behavior may differ by the exposure to different pension schemes, earlier work also show that it varies across different demographic groups. Karlsson and Olsson (2012) find that the level of education has a positive effect on retirement age at the aggregate level; the mean exit age ranges

[1] The 2003 Employment Protection Act (Anstallningsskydd) increased the compulsory retirement age from 65 to 67.

from 61.7 for those only finished compulsory school (Frgymnasial utbildning) to 65 for the post-graduates (Forskarutbildning) in Sweden. Similar effect was found in micro econometric study on the US workers. Berkovec and Stern (1991) show that, regardless of using a static or dynamic model, one more year of schooling decreases the value of retirement. They also show that poor health increases the value of retirement. This finding is corroborated with a study of the retirement behavior of the Dutch old-age workers. Heyma (2004) finds that older Dutch workers with poor health are more likely to enter into early retirement as well as enrol in disability pension. To this end, I further contribute to the literature by examining the cohort-specific behavioral differences by gender, health and education.

The remainder of the paper is organized as follows. The first section briefly introduces the history of the Swedish pension system, followed by illustrating the structural dynamic programming model of retirement. Data source and the estimation procedures are stated in the section of data and method. And results are then reported and discussed in the succeeding section. Finally it concludes by summarizing some key findings.

A Brief History of the Swedish Pension System

The first Swedish pension system was introduced in 1913. It was triggered by the failure of private inter-generational transfer during the industrialization era in Sweden. By then, the share of the population engaged in industry, trade and communication increased dramatically, while decreased considerably in agriculture sector. In the mean time, large scale of labor migrate from countryside to town. This left the remaining old dependants without support, and thus the parliament established a general old age insurance plan. The 1913s pension was the first one in the world covering all citizens regardless of occupation (Palme and Svensson, 1997). However, it meant only minor. The benefit was approximately 11 percent of the earnings of a factory worker and merely a third of the subsistence minimum (Bengtsson and Fridlizius, 1994). By 1930, the replacement rate was still lower than 20 percent (Kruse, 2010).

The Defined-Benefit Pay-As-You-Go (DB PAYG) was introduced in 1960, which supplemented the flat rate basic pension. The pension benefit started to grow continuously. Meanwhile, old-age dependency ratio increased leading to considerable pressure on the pension system. The pressure was fortunately offset by the rapid economic growth at the same time. Over the 1970s, the real wage fell, yet benefit for retirees remained unchanged. Therefore, up until 1980, the overall pension replacement rate reached at over 40 percent of an average industrial workers wage, approximately. Such weak connection between benefit and economic/demographic changes led to unequal distribution of income across generations. As a result, nearly half of the bank savings in the 1990s are owned by those who aged 65 years or older (Bengtsson and Fridlizius, 1994).

A deep recession in the 1990s induced the contribution base shrink by around 10 percent, which stimulates the need for pension reform. Consequently, in June 1994, Swedish parliament passed legislation and replaced the Defined-Benefit Pay-As-You-Go (DB PAYG) with a new system comprising two main pillars. The first pillar is Notional Defined-Contribution Pay-As-You-Go (NDC PAYG or Income Pension) and the second pillar is funded with privately managed individual accounts (Premium Pension). It is, however, supplemented with a guarantee pension at age 65 for persons with low lifetime earnings (Palmer, 2000).

The total contribution rate is 18.5 percent of the pensionable income after the reform. 16 percent of the pensionable income goes to the 1st pillar, NDC PAYG, and 2.5 percent goes to the 2nd funded pillar. The ceiling of the pensionable income is approximately 3000 euro per month. The split aims to create a funded component (i.e. the 2nd pillar), in which individuals can invest their accumulated capital into around 800 privately managed funds at their own discretion, and provide a portfolio mixing economic and financial returns. Individuals can choose the age of retirement and entitle for earning-related pension from age 61 onwards (Kruse, 2008).

There was a gradual transition from the DB PAYG to NDC PAYG. The 1938 cohort received one-fifth pension calculated based on the new rule, and four-fifths based on the old rule. These proportions changed by 5 percent per year for each successive cohort up to those born in 1953. From the 1954 cohort onwards, benefits are accounted by a complete conversion of the accumulated pension credits from the old system into the new system (Palmer, 2000; Settergren, 2001).

Even though the reform did not necessarily lead to substantive benefit cuts, yet it did lose the generosity, to some extent. The amount of the pension income is substantially dependent on the contribution history as well as the remaining life expectancy for each cohort. That is to say that persons with higher lifetime earnings would receive a proportionally higher benefit (Palmer, 2000). In another word, the monthly pension benefit increases with the age of retirement. Hence it provides income incentive for postponing of retirement.

The Retirement Model

Retirement behavior has been a major concern for both economists and politicians over recent decades mainly due to the ongoing process of population ageing in many developed as well as developing countries. The topic has also been increasingly popular within the econometric society. Vast retirement models have been developed ranging from static to dynamic as well as from structural to non-structural. Earlier studies showed that dynamic models better represent the forward looking behavior of individual workers and have stronger predictive power comparing to static models (Lumsdaine et al., 1990; Berkovec and Stern, 1991; Heyma, 2004; Stock and Wise, 1990). Some also argued that structural models are less restrictive than non-structural models. This is because non-structural dynamic models, e.g. Heckman and Macurdy (1980); MaCurdy (1981), assumed wage schedule over the lifecycle for each individual is fixed and independent of participation, i.e. a year increase of working experience would have no effect on wage. Structural models, on the other hand, incorporates various effects on the dynamic development of wage, e.g. age, experience, uncertainty as well as "job match", and therefore relax the assumption of fixed pay schedule over the life-cycle (Berkovec and Stern, 1991; Lazear, 1987; Hurd, 1990; Quinn and Burkhauser, 1990). For this, the retirement model in this paper is not only dynamic, but also structural in the sense that the anticipated wage schedule is conditional on individual participation decision, i.e. years of working. The structure of retirement model is succinctly stated in what follows.

The model assumes that each agent confronts the choice set, retire or continue working , at each period. Retirement is assumed to be an absorbing state, i.e. there is no possibility of re-entry to the labor market after the retirement decision has been made. The job episodes for each agent is modelled over a finite horizon and discrete time. Individual preferences are assumed to be represented by a constant relative risk aversion utility function (CRRA), in which utility is only derived from goods consumption. Thus the utility function can be written as in (1).

$$U(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma} \tag{1}$$

where, let γ be a risk aversion parameter and c_t be the goods consumption.

At each period t, individual maximizes their utility subject to a dynamic budget constrain (2).

$$c_t \leq a_t r_t + y_t(1 - D_t) + b_t D_t \tag{2}$$

where, a_t is the asset stock, r_t is the interest rate, y_t is the net labor income, b_t is the net pension income, and D_t is an indicator variable, $D_t = 1$ if individual is retired and $D_t = 0$, otherwise.

Assuming the rental price for capital is extremely low, i.e. $r_t \approx 0$, the consumption at each period is, therefore, financed through either labor income or pension benefit. Furthermore, the zero interest rate also presumes that rational consumer shall have no saving motive. This implies the equality between consumption and budget constraint in (2). Given these assumptions, the consumption variable in (1) can be replaced by y_t and b_t with respect to the retirment decision D_t. Hence, the utility function in (1) can then be re-written as,

$$U(y_t) = \frac{(y_t)^{1-\gamma}}{1-\gamma} \quad i.i.f \quad D_t = 0 \tag{3}$$

$$U(b_t) = \frac{(\alpha b_t)^{1-\gamma}}{1-\gamma} \quad i.i.f \quad D_t = 1 \tag{4}$$

where, let α be the utility weight attached to retirement benefit while not working relative to the utility weight attached to labor income while working.

The presence of α in (4) is inspired by Lumsdaine et al. (1990); Stock and Wise (1990). This parameter is to recognize the difference between the utility associated with a dollar of income accompanied by working and the utility associated with a dollar of income while retired.

Static Choice Probelm

In the static scenario, individuals are assumed to be myopic. Therefore, in deciding whether to retire or not, only the current utility derived either from current entitled pension or labor earnings is taken into account. In order for a rational agent to retire, a necessary condition is that the utility of retire is greater than or, at least, equal to the utility of continue working. This argument implies that the ratio of (4) to (3) is greater than or equal to 1. When this ratio strictly equals to 1, it can be interpreted as a special case - the minimum requirement for the retirement transition to occur. That is to say that the utility derived from pension benefit is indifferent than that derived from labor income. This minimum requirement condition can be derived by setting the ratio of (4) to (3) equals to 1, that is,

$$\left\{\alpha\frac{(b_t)}{(y_t)}\right\}^{1-\gamma} = 1 \tag{5}$$

The term b_t/y_t in (5) is essentially the replacement rate. Taking the power of $1/(1-\gamma)$ on both sides of (5), it yields a relation between the replacement rate and the parameter α. Since (5) refers to the minimum requirement for retirement, the term b_t/y_t can then be interpreted as the required rate of replacement for retirement and denoted by R^*. Hence the required rate of replacement for retirement, R^*, can be expressed as,

$$R^* = \frac{1}{\alpha} \tag{6}$$

The interpretation of R^* is straightforward. When α equals to 1, R^* will be 1 as well, and thus it implies that an individual would require the amount of pension benefits exactly equals to the amount of labor income so as to retire. Should $\alpha < 1$, workers would require the retirement income that is R^* times of the current wage. And if $\alpha > 1$, the required pension is R^* percent of the labor earnings. The greater the α is, the smaller the R^* will be, and thereby the higher the propensity of retirement transition is.

Replacement rate is a common indicator for the level of retirement income relative to labor earnings. Each person covered by the Swedish Pension System receives an individual report - "Orange Envelop" - on the prospective pension benefits as well as the replacement rate. The projected entitlements and rates can vary by the age at retirement. Hence information on the income during retirement is foreseeable and updated once a year for individuals. The Swedish Pension Board (Pensions Myndigheten) also projects the average replacement rates for different cohorts and publishes the results in its annual report, the "Orange Report". The calculation typically sets age 65 as a benchmark, and computes the pension level at age 65 as a share of average income over age 60-64. Nevertheless, this replacement rate per se is meaningless for evaluating the pension system. Firstly, retirement age differs by cohorts, as shown by Karlsson and Olsson (2012). Secondly, the meaning of replacement rates might also vary to different individuals, e.g. it might be sufficient for certain group, yet insufficient for others. This may, in turn, translate into behavioral differences in retirement transition, which is important for evaluating the sustainability of the pension system. To understand why cohort retires at different ages, given the differentials in replacement rates, the required rate of replacement is useful. Since it gives the replacement level necessary for an individual to retire. And its variation together with the variation in replacement rate explains why certain people retire later. The basic assumption is that workers choose to retire if, and only if, the projected replacement rate exceeds the required rate of replacement. From a policy perspective, if a government attempts to increase retirement age so as to sustain the pension system, adjusting the replacement rate in accordance with the required rate of replacement might be a feasible instrument.

Dynamic Choice Probelm

To solve the dynamic problem of either retire or continue working, individuals are assumed to be forward looking. That is not only the current utility, but also the future utility flows are taken into account when making the choice. Hence, at each time period t, every individual confronts a choice set: 1) retire and derive utility from current and future pension benefits, or 2) work and derive utility from labor earnings for the current period, and leave the retirement option open for the next period. Hence, at each period, individual maximizes a value function expressed as,

$$V_t = \max \left\{ U(y_t) + \beta E(V_{t+1}) + \epsilon_{1t} \ , \ \sum_{t}^{T} \beta^{T-t} U(b_t) + \epsilon_{2t} \right\} \tag{7}$$

where, β is a discount factor, T is the life expectancy, and $E(V_{t+1})$ is the expected option value for next period. ϵ_{1t} and ϵ_{2t} are the errors assumed to be i.i.d.

The term $E(V_{t+1})$ is computationally complex as it can only be solved numerically. However, following Berkovec and Stern (1991), the analytical solution exists i.i.f the $\epsilon' s$ are assumed to follow i.i.d. extreme value distribution. The dynamic programming of the future expected value can then be solved by backward recursive computation using the value function in (8).

$$E(V_{t+1}) = \tau \left\{ u + \ln \left\{ \exp\left(\frac{U(y_{t+1}) + \beta V_{t+2}}{\tau}\right) + \exp\left(\frac{\sum_{t+1}^{T} \beta^{T-t+1} U(b_{t+1})}{\tau}\right) \right\} \right\} \tag{8}$$

where, u is a Euler constant and τ is the scale parameter of the extreme value distribution.

For (8), the terminal condition is defined as the expected future value at the highest possible age for retirement. I assume the latest retire age is 70 years old. Hence, for $t = S = 70$, the expected value at the terminal condition is computed using (9).

$$E(V_S) = \tau \left\{ u + \ln \left\{ 1 + \exp\left(\frac{\sum_{S}^{T} \beta^{T-S} U(b_S)}{\tau}\right) \right\} \right\} \tag{9}$$

The assumption that $\epsilon' s$ are i.i.d. draws from the extreme value distribution gives the closed form expression for the probability of working and the probability of retire, respectively, as in (10) and (11),

$$\begin{aligned} Pr(D_t = 0) &= Pr\left(U(y_t) + \beta E(V_{t+1}) + \epsilon_{1t} > \sum_{t}^{T} \beta^{T-t} U(b_t) + \epsilon_{2t}\right) \\ &= \frac{\exp\left(\frac{U(y_t) + \beta E(V_{t+1})}{\tau}\right)}{\exp\left(\frac{U(y_t) + \beta E(V_{t+1})}{\tau}\right) + \exp\left(\frac{\sum_{t}^{T} \beta^{T-t} U(b_t)}{\tau}\right)} \end{aligned} \tag{10}$$

$$Pr(D_t = 1) = Pr\left(U(y_t) + \beta E(V_{t+1}) + \epsilon_{1t} \leq \sum_t^T \beta^{T-t} U(b_t) + \epsilon_{2t}\right)$$
$$= \frac{\exp\left(\frac{\sum_t^T \beta^{T-t} U(b_t)}{\tau}\right)}{\exp\left(\frac{U(y_t) + \beta E(V_{t+1})}{\tau}\right) + \exp\left(\frac{\sum_t^T \beta^{T-t} U(b_t)}{\tau}\right)} \tag{11}$$

The inequality between $U(y_t) + \beta E(V_{t+1}) + \epsilon_{1t}$ and $\sum_t^T \beta^{T-t} U(b_t) + \epsilon_{2t}$ in (11) implies that, for a rational individual to retire, the sum of the current and the discounted future utility of pension must greater than or equal to the sum of the current utility of labor income and the expected option value. And the minimum requirement condition for the retirement transition is these two terms equal to each other. This condition is no different than that in the static scenario discussed formerly. Hence, the interpretation for R^* remains the same. Nevertheless, the value of α as well as R^* can differ from the static case simply because the future utility flows are taken into account when agent confronts the retirement decision. The utility of retirement would increase along with worker ages, while the utility of continue employment might decrease. Thereby, α is expected to be larger (or equivalently, R^* is smaller) in the dynamic framework than in the static setting.

The magnitude of the differences in α or R^* between the static and dynamic case is determined by the discounting factor β. In another word, the differences in estimates are dependent on the degree of the forward-looking behavior of each individual. A large discounting factor, or strong foresight, implies that the future utility flows matters when deciding whether to retire or not today. And thus the estimates of α or R^* would be considerably differ by the static and dynamic assumption. Conversely, a small discounting factor, or individuals are myopic, implies that the future utility flows means little to the current retirement decision. And therefore the estimates in the static and dynamic setting would be close to each other.

Data

Data for the analysis in this paper is sourced from the working life histories in the Survey of Health, Ageing and Retirement in Europe (SHARELIFE). It provides information on different episodes over the entire life course for each observed person, such as years in education, spells of working, unemployment, retirement, etc. The dataset is a panel in nature, but constructed through retrospective survey.

To address the research question: whether the retirement behavior of the Swedish notch babies differs than their older counterpart, I only use the Swedish sub-sample that contains 1893 individuals, of which 848 males and 1045 females. Ideally, all the cohorts should be used for the analysis so that the retirement behavior for those born in 1937 or earlier can be compared with the notch babies, i.e. born in 1938 or later. However several concerns arise.

First of all, including very old cohorts might introduce selection bias. For instance, if we include those born in 1930, by the year of interview (2008/2009), they are already nearly 80 years old. This might imply that these participants are healthier than non-participants in the same cohort and therefore more likely to survive to the survey date. It could, in turn, assure that their labor market outcome is better than the others within the cohort. Moreover, due to the nature of the retrospective survey, old participants might be more likely associated with recalling bias. Therefore, the "very old" generations are excluded in the sample for the analysis.

Secondly, retirement, as an outcome during the later life, might be largely determined by the early experience over the life course. Such experience might greatly differ between those born in 1930 and 1940 as the macro economic and environmental conditions they exposed to are different. To examine the difference across large cohort span requires controlling for such macro conditions. This, however, is a burdensome task in this study due to little information in this regard is given in the data. Hence, to ensure the comparability of the cohorts and their exposures to the similar macro environment, the sample is restricted to those born in 1937 or later.

Finally, the retrospective yearly information on whether working or retired for each individual ends by year 2009 in SHARELIFE. And, as discussed in the proceeding section, the terminal condition for the backward recursive computation of the expected value function is assumed be at age 70, thus those born after year 1938 has to be excluded from the sample. To do so is to assure the same length of exposure to retirement risk, i.e. age span 60-70, across cohorts. Hence, the sample is restricted to those born in 1937 and 1938, of which the older cohort was unaffected by the 1994 pension reform, whereas the younger cohort is the notch babies.

The discrete choices of work or retirement overtime for all individuals are the outcome variable in this study. And the covariates used are: after-tax monthly wage at the end of main job, after-tax monthly work income at the end of main job, after-tax monthly first pension benefit, gender, year of birth, age, and years in full-time education. Furthermore, SHARELIFE does not provide health record for each individual over time. Nonetheless these information is obtained by linking the individual in SHARELIFE to the same respondent in the third wave of SHARE, where the self-evaluated general health status is available. The method of dealing with each of the variables are discussed in the method section. Each individual is assumed to expose to the risk of retirement over the age span 60-70, and once the retirement is taken place, the individual is censored from the data. Given this censoring mechanism, the final sample contains 78 individuals and 452 observed person-years.

Method

The retirement model formerly stated comprises only two covariates, the labor income if working and benefits if retired. These variables, however, are not directly available in SHARELIFE. And therefore the retirement model is estimated by a two-stage-procedure.

In the 1^{st} stage, the random effects model using Hierarchical Bayesian Estimation is applied to estimate a wage function. The estimation is implemented by Bayesian Inference Using Gibbs Sampling (WinBUGS). The model is stated as following,

$$y_i \sim N(\alpha_i, \sigma^2)$$

$$\alpha_i \sim N(\bar{\alpha_i}, \lambda^2)$$

$$\begin{aligned} \bar{\alpha_i} = \alpha_0 + \alpha_1 age_i + \alpha_2 exper_i + \alpha_3 exper_i^2 + \alpha_4 educ_i + \alpha_5 male_i \\ + \alpha_6 goodhealth_i + \alpha_7 male_i * exper_i \end{aligned} \tag{12}$$

where, i denotes observed individual, y is the labor income, α_i is the random coefficient for each individual, $\bar{\alpha_i}$ is the mean of the random coefficient α_i, σ^2 and λ^2 are the variance of observed income across individuals and the variance of the random coefficient α_i, respectively, which are both assumed to follow inverse gamma distribution, i.e. $IG(.01, .01)$, and α_0 to α_7 are the coefficients for all the covariates in the wage equation, which are assumed to be $N(0, 0.0001)$.

The data used at this stage is the sum of the last wage at the end of main job and the last work income at the end of the main job in SHARELIFE. All the values are nominal value of local currency at time. Therefore, they are all converted to 2003 price level by using the price indices 1830-2010 from Statistics Sweden. All the other covariates in the wage equation are: age, years of working experience, education level, gender, health, and a interaction term of gender and experience[2].

The estimated random coefficients together with the observed information on age, education, experience, gender and health status in SHARELIFE are then used to impute the individual labor income streams over the age span 60-70. The model is structural in the sense that the expected wage growth curve for workers will be not only dependent on age, but also conditional on the work participation.

SHARELIFE collects information on the first monthly after-tax pension benefit at the nominal values of local currency at time. These values are converted to the 2003 price level following the same procedure as labor income. However the entitled benefits are assumed be constant over the ages between 60 and 70. That is to say that the model only allows for cross-individual variation in pension income, and the variation is held constant over time.

In the 2^{nd} stage, the multinomial logit model is applied to fit the individual discrete choices of work or retirement over age 60-70 using the Hierarchical Bayesian Estimation (HB). The main reason to apply HB estimation other than the Maximum Simulated Likelihood (MSL) is that the sample size is small, merely 78 individuals. This might lead to the case of asymptotic properties not being fully exhibited. As argued by Train (2009), unlike the classical perspective requires the asymptotic assumption of the sampling distribution, which might not be fulfilled when the sample size is insufficient, the posterior distribution in the Bayesian approach contains the information with any sample size, and therefore suitable for

[2]Years of working experience is the sum of all working spells for each individual, education is a dummy variable (equal to 1 if individual completed 15 years in full-time education, which is equivalent to finish tertiary education, and 0 otherwise), male is a gender dummy (equal to 1 if being male, and 0 otherwise), good health is a dummy of self-reported general health status (equal to 1 if it is excellent or very good or good, 0 if fair or poor)

small sample inference. The HB estimation procedure is stated as following. From (10) and (11), the logit for each individual i over time can be formed as,

$$L(D_{i,t} \mid \gamma_i, \alpha_i, \beta_i) = \prod_t \left\{ \frac{\exp\left(\frac{U(y_{i,t}) + \beta E(V_{i,t+1})}{\tau}\right)}{\exp\left(\frac{U(y_{i,t}) + \beta E(V_{i,t+1})}{\tau}\right) + \exp\left(\frac{\sum_t^T \beta^{T-t} U(b_{i,t})}{\tau}\right)} \right\}^{1-D_{i,t}} \times \left\{ \frac{\exp\left(\frac{\sum_t^T \beta^{T-t} U(b_{i,t})}{\tau}\right)}{\exp\left(\frac{U(y_{i,t}) + \beta E(V_{i,t+1})}{\tau}\right) + \exp\left(\frac{\sum_t^T \beta^{T-t} U(b_{i,t})}{\tau}\right)} \right\}^{D_{i,t}} \tag{13}$$

Let θ_i be a vector of individual random parameters $\gamma_i, \alpha_i, \beta_i\ \forall i$ and $\theta_i \sim N(\Theta, W)$, where, Θ is the vector of population level parameters $\bar{\gamma}, \bar{\alpha}, \bar{\beta}$ and W is the variance-covariance matrix for the population level parameters. The prior is thereby $p(\Theta, W) = p(\Theta)p(W)$, where, $p(\Theta) \sim N(\Theta_0, S_0)$ with extremely large variance, $p(W) \sim IW(3, I)$, which is Inverse Wishart Distribution with 3 degree of freedom and a 3-dimensional identity matrix. The three conditional posteriors are therefore as following,

$$\mathbf{P}(\Theta \mid W, \theta_i\ \forall i) \sim N(\bar{\theta}, W/N) \tag{14}$$

where, N is number of observed individuals, and $\bar{\theta} = \sum_{i=1}^{N} \frac{\theta_i}{N}$

$$\mathbf{P}(W \mid \Theta, \theta_i \ \forall i) \sim IW\left(3 + N, \frac{3 \times I + N \times \bar{S}}{3 + N}\right) \tag{15}$$

where, $\bar{S} = \sum_{i=1}^{N} \frac{(\theta_i - \Theta)(\theta_i - \Theta)'}{N}$

$$\mathbf{P}(\theta_i \mid \Theta, W, D_{i,t}) \propto \prod_t L(D_{i,t} \mid \gamma_i, \alpha_i, \beta_i) \ \phi(\theta_i \mid \Theta, W) \ \forall i \tag{16}$$

where, $\phi(\theta_i \mid \Theta, W)$ is the normal density of $\theta_i's$ conditional on hyper-parameters Θ and W.

All the parameter estimates are obtained from the above three conditional posteriors by the Gibbs Sampling. (14) and (15) give the hyper-parameters and their corresponding variance-covariance, and (16) provides the individual parameters for each i. For both static and dynamic models, the number of iterations for Monte Carlo Markov Chain is set to 40000, of which the first 20000 iterations are burn-in.

To examine the cohort, education, gender, and health differences in retirement behavior, the individual-level random coefficients are analysed, which are sampled from the conditional posterior (16). As discussed in the retirement model, the required rate of replacement, R^*, is the primary interest in this paper since it measures the retirement propensity. And the R^* is the inverse of α, shown in (6). Therefore the differences in retirement behavior is examined by regressing the individual-level parameter R_i^* on education, gender, and health status for each of the two cohorts, respectively. The results are reported in the succeeding section.

Results and Discussion

Wage Estimation and Imputation

Table 1 reports the parameter estimates for the wage equation (12) using Hierarchical Bayesian Estimation. The wage rates are transformed into logarithm scale. The estimated coefficients suggest that wage drops at a rate of 8 percent per year between age 60 and 70. Completing tertiary education (year of schooling greater than or equal to 15) leads to nearly 23 percent wage premium. Male workers earn more than twice higher than female workers, whereas the return to an extra year of working experience is lower for male than female, as indicated by the interaction term of male and experience. And good health condition seems matter considerably in earning outcomes, over 15 percent higher on average than workers with poor health.

The estimated coefficients in Table 1 are then used for imputing the expected wage stream for each individual over the age 60-70. For each individual, $\bar{\alpha}_i$ is first obtained by calibrating the parameters in Table 1 to (12). Then the imputation proceeds by evaluating the two integrals in (17). Both integrals are approximated by taking the means of 10000 draws for α_i and y_i from the two corresponding normal density $\phi(\bar{\alpha}_i, \lambda^2)$ and $\phi(E[\alpha_i], \sigma^2)$, respectively.

$$\begin{aligned} E[\alpha_i] &= \int \bar{\alpha}_i \; \phi(\bar{\alpha}_i, \lambda^2) \; d\bar{\alpha}_i \\ E[y_i] &= \int E[\alpha_i] \; \phi(E[\alpha_i], \sigma^2) \; dE[\alpha_i] \end{aligned} \tag{17}$$

Population-level Parameter Estimates

Table 2 reports the population-level parameters and their standard errors estimated by (14) and (15). In estimation, the scale parameter τ in (13) is set to one due to the non-convergence of its Markov Chain. The static model estimates the risk aversion parameter γ and the parameter α, which is for calculating the required rate of replacement R^*. The dynamic model, on the other hand, has one

Table 1: Estimation of Wage Equation (12)

Coef	Est.	SE
$Constant$	17.933	(2.438)
Age	-0.080	(0.026)
$Exper$	-0.069	(0.077)
$Exper^2$	0.001	(0.001)
$Educ$	0.228	(0.115)
$Male$	2.111	(0.724)
$Health$	0.151	(0.098)
$Male * Exper$	-0.040	(0.017)
λ^2	0.237	(0.086)
σ^2	0.239	(0.085)
obs	72	
chains	3	
iterations	20000	

more parameter to be estimated, that is the discounting factor, β. It is important to stress that the coefficients and standard errors reported in Table 2 are the means of each element in Θ in (14) and on the diagonal of W in (15) over the last 20000 iterations[3]. Hence, the interpretation of these estimates is no different than the point estimates and standard errors in the standard maximum likelihood estimation. Nevertheless, it does not necessarily mean that the value of the posterior mean would be identical to the point estimates in the maximum likelihood estimation because they could differ if the sample size is insufficient for the asymptotic convergence (Train, 2009).

The two models yield very similar estimates in both γ and α. This is mainly because the estimated discounting factor β is as low as 0.3. It suggests that, on average, individuals weigh the current utility much greater than (more than three times of) the future utility when deciding whether to retire today. Based on the risk aversion parameter γ, the inter-temporal elasticity of substitution is fairly strong, 1.326 in the static model and 1.277 in the dynamic model. It implies that

[3]The standard errors are computed by taking the square root of the diagonal of the variance-covariance matrix, W.

the consumption growth with respect to changes in real interest rate is substantially larger than unity.

The α coefficients indicate that the utility weight on the non-labor income is over 30 percent higher than the income accompanied by work. Using (6), the required rate of replacement, R^*, is 0.76 and 0.74 in both models, respectively. It can be interpreted as the necessary level of pension entitlements in order for an individual to retire is 76 percent (in the static model) and 74 percent (in the dynamic model) of the labor earnings, respectively. Furthermore, R^*'s also suggest that retirement propensity is slightly higher in the dynamic model than the static model. This is because the dynamic model takes all future utility flows into account, and the utility of retirement relative to continued working increases with age.

Table 2: Population Level Parameter Estimates of the Retirement Model

	Static		Dynamic	
Coef.	Est.	SE	Est.	SE
γ	0.754	(0.140)	0.783	(0.134)
α	1.316	(0.225)	1.358	(0.246)
β			0.307	(0.212)
obs	78		78	
Iterations	40000		40000	
Burn-ins	20000		20000	
Log-likelihood	-112		-123	

Cohort Differences in Retirement Propensity

Table 3 summarizes the individual-level random coefficients R_i^* for the 1937 cohort who were unaffected by the 1994 pension reform and the notch babies, respectively. Similar to the population-level estimates, the difference between the static and dynamic model is fairly small. As discussed in the proceeding section, this is attributable to the small discounting factor estimated in the dynamic model. And the weight on the current utility is much greater than on the future utility when it comes to the retirement decision. Nevertheless, the means and quantiles of R_i^* are slightly smaller in the dynamic model for both cohorts. This is mainly due to the fact that the utility of retirement relative to continued working increases with age. And such increases is only captured in the dynamic model since future utility flows were taken into account for the retirement decision today.

Overall, the cohort differences in the required rate of replacement are negligible. Only the quantiles show that R_i^* is slightly smaller, yet the variance of R_i^* is greater for the notch babies. Generally speaking, the summary statistics imply that the 1994 pension reform had little impact on the retirement behavior, while it induced more heterogeneous preferences on work-retire choices. To examine whether this conclusion can be held, the cohort-specific difference in R_i^* with respect to educational attainment, gender, and health status is analysed in the succeeding section.

Table 3: Summary of Individual-level R_i^* by Cohort

	Static		Dynamic	
Cohort	1937	1938	1937	1938
1^{st} Quantile	0.6654	0.6319	0.6477	0.6098
Median	0.7286	0.6996	0.7267	0.6857
Mean	0.8123	0.8199	0.7955	0.7982
3^{rd} Quantile	0.8948	0.8528	0.8777	0.8286
Variance	0.0547	0.0830	0.0594	0.0852

Cohort-specific Heterogeneity

Table (4) reports the results for the cohort-specific regression of R_i^* on a set of covariates using Ordinary Least Square. Overall, all coefficients are statistically insignificant, except all the constants and the education coefficient for the 1937 cohort.

The magnitude and standard errors of the education coefficients for the older cohort are identical in both static and dynamic models. The education parameter estimates indicate that within the generation who were unaffected by the pension reform, the required rate of replacement is nearly 0.2 higher for those completed tertiary education. This relation, however, was reversed among the notch babies. The education coefficients become negative for the 1938 cohort, although statistically insignificant.

Gender differences in R_i^* was small for the 1937 cohort, in terms of both the size and the significance. Nonetheless, male born in 1938 was more reluctant to retirement than female, shown by the male coefficient, 0.06 and 0.09 in the static and dynamic model, respectively.

The health coefficients appear small and far from significance in all four regressions. However it is noteworthy that the negative coefficients imply that those with good health status even have less incentives for prolong working life, despite their average labor earnings are 15 percent higher than the unhealthy workers (shown in Table 1).

To date, when controlling for education, gender, and health status, retirement behavior at the individual-level varies considerably across cohorts. Interpreted as required rate of replacement, tertiary education has large and significant effect on R*, while gender and health have no impact for those unaffected by the pension reform, ceteris paribus. Nevertheless, among the Notch Babies, the positive effect of higher education on R^* is reversed, while gender differences become larger, ceteris paribus. Such cohort differences are identical in both static and dynamic models.

Conclusion

The Swedish pension system underwent a major structural change during the late 1990s. By June 1994, the parliament passed legislation and replaced DB PAYG with a new pension system comprising of two pillars: NDC PAYG and Premium Pension. The oldest notch babies, the first generation who were effectively af-

Table 4: Heterogeneity of R_i^* by Cohort

Cohort	1937		1938	
Coef.	Est.	*SE*	Est.	*SE*
	R_i^* in Static Model			
Intercept	0.768***	(0.091)	0.815***	(0.125)
Educ	0.168**	(0.078)	-0.003	(0.107)
Male	-0.005	(0.079)	0.066	(0.096)
Health	-0.045	(0.079)	-0.036	(0.132)
obs	39		39	
DF	35		35	
R^2	0.127		0.016	
	R_i^* in Dynamic Model			
Intercept	0.707***	(0.095)	0.770***	(0.126)
Educ	0.168**	(0.082)	-0.039	(0.108)
Male	0.030	(0.083)	0.091	(0.096)
Health	-0.013	(0.083)	-0.010	(0.132)
obs	39		39	
DF	35		35	
R^2	0.110		0.030	

Significance Codes: *** 0.01, ** 0.05, * 0.1

fected by this reform, are those born in 1938. The primary question this paper addresses is whether the retirement behavior of the Notch Babies is different than their older counterpart, i.e. the 1937 cohort that was unaffected by the reform. Some aggregate statistics show that the mean age at job market exit is higher for the notch babies relative to the 1937 cohort; therefore, the second purpose of this paper is to verify whether the macro evidence can be supported by the findings from micro-level analysis.

The retirement model in this study is similar to earlier models developed by Lumsdaine et al. (1990); Berkovec and Stern (1991); Heyma (2004); Stock and Wise (1990). Unlike many earlier studies using maximum likelihood, pseudo maximum likelihood, and/or the method of simulated moments for estimation, I, however, apply the Hierarchical Bayesian Estimation to estimate the wage function, impute the missing wage data, and fit the discrete choice model. This is because the sample size for the analysis is rather small, merely 78 individuals. The insufficient sample size prevents me from fulfilling the asymptotic assumption of the sample distribution as well as from using any Maximum Likelihood Estimation. The posterior distribution in the Bayesian approach is suitable for small sample inference, and therefore is applicable to survey data, e.g. SHARELIFE, where observations are not redundant.

The model is estimated under two assumption on the forward-looking behavior of individuals. The static model assumes individuals are myopic, for which, the discounting factor is set to zero, while the dynamic model represents the perfect foresight assumption, where the discounting factor β is estimated. Both models yield very similar results mainly because the estimated β is low. This suggests that individuals weigh the current utility much higher than the future utility flows when deciding whether to retire today.

At the population level, the results show that the population at age 60-70 are with strong inter-temporal elasticity of substitution, of which the consumption growth with respect to changes in real interest rate is substantially greater than unity. The estimated required rate of replacement, R^*, is slightly lower in the dynamic model. This is because all of the future expected utility is taken into account by the dynamic model, and the ratio of the utility of retirement to working increases with age. R^* displays no difference between the 1937 cohort and the notch babies, and thus lends no support to the aggregate pattern that the mean retirement age increased from the 1937 cohorts onwards. Nonetheless, the analysis of the individual-level parameters show that the reform encouraged those with completed tertiary education to retire early, while simultaneously prompting male workers to prolong working life. Meanwhile, there is no difference between healthy and unhealthy workers in retirement behavior across cohorts, despite the fact that the average labor earnings are 15 percent higher for those with good health status.

To date, the estimates of preference parameters in both dynamic and static models are in line with the findings of Stock and Wise (1990). The cohort differences in retirement behavior is congruent with the aggregate pattern only among the male population, yet not among the educated. Finally, the empirical findings contradict those of Berkovec and Stern (1991) and Heyma (2004) who showed that good health has large and adverse effect on retirement probability.

The retirement model in this paper is simple, but with two strong assumptions: zero interest rate and only two discrete choices, work or retire. The former is mainly due to the unavailability of wealth information. The latter is because alternative working status (e.g. partial retirement, job switches, and laid off, etc) as well as the associated cost and benefit of each status are not directly identifiable. Hence, future research should focus on the wealth effects on retirement and developing multi-choice models; however, this requires datasets with a larger sample and more information. Furthermore, the findings reported in this paper could be flawed for three reasons.

Firstly, the number of observed persons is small. This prevents me from controlling for more explanatory variables when examining the cohort-specific differences, e.g. occupation and industry, etc. The small sample size might also contribute to the lack of congruence between the micro evidence and the macro emergence if the sample is not representative.

Secondly, the retrospective survey on life history could potentially introduce measurement error, such as recalling biases and/or selection biases. Whether the errors are systematic or random, if there is any, deserves special attention before drawing any substantive policy implication.

Finally, the health information is self-reported. The reliability of such self-evaluation remains unclear. Hence, the interpretation of the weak health effects on retirement needs to be treated with caution. Nevertheless, the aforementioned problems could be resolved by comparing the results with similar study based on other data source, preferably with larger sample and more objective measurements. As a result, the priority in future research will be to conduct similar analysis using the Swedish Income Registers in order to validate the results reported in this paper.

References

Bengtsson, T. and G. Fridlizius (1994). *The Family, the Market and the State*, Chapter Public Intergenerational Transfers as an Old-Age Pension Systems: A Historical Interlude?, pp. 198–215. Oxford University Press.

Berkovec, J. and S. Stern (1991, January). Job Exit Behavior of Older Men. *Econometrica 59*(1), 189–210.

Heckman, J. J. and T. E. Macurdy (1980). A life cycle model of female labour supply. *The Review of Economic Studies 47*(1), pp. 47–74.

Heyma, A. (2004). A structural dynamic analysis of retirement behaviour in the Netherlands. *Journal of Applied Econometrics 19*(6), 739–759.

Hurd, M. D. (1990). Research on the elderly: Economic status, retirement, and consumption and saving. *Journal of Economic Literature 28*(2), pp. 565–637.

Karlsson, H. and H. Olsson (2012). Medelpensioneringsalder och uttradesalder 2011. *Pensions Myndigheten.*

Karlstrom, A., M. Palme, and I. Svensson (2004). A dynamic programming approach to model the retirement behaviour of blue-collar workers in Sweden. *Journal of Applied Econometrics 19*(6), 795–807.

Krueger, A. B. and J. Pischke (1992). The effect of social security on labor supply: A cohort analysis of the notch generation. *Journal of Labor Economics 10*, 412–37.

Kruse, A. (2008). Expected effective retirement age in the nordic countries. *Working Paper in Social Insurance 3.*

Kruse, A. (2010). A stable pension system: The eighth wonder. In T. Bengtsson (Ed.), *Population Ageing - A Threat to the Welfare State?*, Demographic Research Monographs, pp. 47–64. Springer Berlin Heidelberg.

Lazear, E. P. (1987). Retirement from the labor force. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbook of Labor Economics*, Chapter 5, pp. 305–355. Elsevier.

Lumsdaine, R. L., J. H. Stock, and D. A. Wise (1990). Three models of retirement: computational complexity versus predictive validity. *National Bureau of Economic Research Working Paper 3558*.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy 89*(6), 1059–85.

Palme, M. and I. Svensson (1997, August). Social Security, Occupational Pensions, and Retirement in Sweden. NBER Working Papers 6137, National Bureau of Economic Research, Inc.

Palmer, E. (2000, June). The swedish pension reform model : framework and issues. Social Protection Discussion Papers 23086, The World Bank.

Quinn, J. F. and R. V. Burkhauser (1990). Work and retirement. *Handbook of aging and the social sciences 3*.

Settergren, O. (2001). *Two Thousand Five Hundred Words on The Swedish Pension Reform*.

Stock, J. H. and D. A. Wise (1990, September). Pensions, the option value of work, and retirement. *Econometrica 58*(5), 1151–80.

Train, K. E. (2009, November). *Discrete Choice Methods with Simulation*. Cambridge University Press.