MEMORANDUM

No 3/2008

Dynamic programming model of health and retirement

Fedor Iskhakov

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 2/08	Atle Seierstad and Sigve D. Stabrun *Discontinuous control systems*
No 1/08	Kjell Arne Brekke, Rolf Golombek and Sverre A. C. Kittelsen *Is electricity more important than natural gas? Partial liberalization of the Western European energy markets*
No 23/07	Geir B. Asheim and Taoyuan Wei *Sectoral Income*
No 22/07	Geir B. Asheim and Seung Han Yoo *Coordinating under incomplete information*
No 21/07	Jo Thori Lind and Halvor Mehlum *With or Without U? The appropriate test for a U shaped relationship*
No 20/07	Ching-to Albert Ma *A Journey for Your Beautiful Mind: Economics Graduate Study and Research*
No 19/07	Simen Markussen *Trade-offs between health and absenteeism in welfare states: striking the balance*
No 18/07	Torbjørn Hægeland, Oddbjørn Raaum and Kjell Gunnar Salvanes *Pennies from heaven - Using exogenous tax variation to identify effects of school resources on pupil achievement*
No 17/07	B. Bratsberg, T. Eriksson, M. Jäntti, R. Naylor, E. Österbacka, O. Raaum and K. Røed *Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries*
No 16/07	Kjell Arne Brekke, Gorm Kipperberg and Karine Nyborg *Reluctant Recyclers: Social Interaction in Responsibility Ascription*

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

Dynamic programming model of health and retirement

Fedor Iskhakov[a]

Abstract: A structural dynamic programming model is applied for modeling labour market transitions among older age workers in Norway in 1992-2003. Special attention is given to early retirement pension and disability pension as two major exit routes from the labour force. Health status is represented by a latent variable reflecting the eligibility for participating in disability programs. Incomplete information maximum likelihood method is used in several stages to facilitate the estimation.

The model is used to investigate the degree of potential substitution of the early retirement and retirement through the disability insurance scheme. Estimates of the structural parameters of the concealed health process allow for forecasting the individual "eligibility" for the disability and thus facilitate the assessment of the potential substitution between the two exit routes from the labour force. Performed policy simulation of the complete elimination of the early retirement program indicates nearly complete return of the otherwise early pensioners back to the labour market.

Keywords: Retirement, health, early retirement, disability, labour market transitions, structural dynamic model, dynamic programming.

JEL: J22, J26, I10, C61

---

[a] The current master thesis was written within the Frisch Center project 1132 (Yrkesaktivitet blant eldre og finansiering av pensjonssystemet). Financial support from the Research Council of Norway is greatly acknowledged.

Ragnar Frisch Centre for Economic Research. Gaustadalleen 21, 0349 Oslo, Norway.

Direct phone: +47 22958837. Email: f.v.iskhakov@frisch.uio.no

1. *Introduction and review of the literature*

Recent trend of increasing life expectancy and earlier withdrawal from the labour market has been witnessed in many European countries including Norway. This trend has been threatening the financial stability of the social security systems of the PAYGO type. Forecasts suggest that in the absence of major structural change in the National Insurance System in Norway (NIS) expenditure on old age pension will increase from 6 to about 15 percent of Mainland Norwegian GDP before year 2050 (Summary of report to Storting nr. 12, 2005). Although the retirement age in Norway is relatively high, there is a generous access to disability benefit, which has been playing an important role in lowering labour force participation among older persons (both in Norway (Røed and Haugen, 2003) and other countries (Bound and Waidmann, 1992)). Introduction of an early retirement program (AFP) in 1989 has contributed further to the reduction of labour force participation by providing strong incentives to stop working when the AFP retirement option is available (Bratberg et al., 2004). Hence, policy changes designed to induce higher labour force participation or at least to slow down the decline should be considered with a particular attention to both the AFP retirement process and the disability retirement as they are the most frequently taken exit routes from the labour market.

The current paper develops a structural dynamic model – the tool best suited for comprehensive policy design – which primarily focuses on these two major exit routes from the labour market. The model is formulated on the individual level and keeps track of individual differences with respect to the retirement alternatives available for each decision maker. Whereas the AFP pension eligibility rules are well established and documented (for description see Appendix, p. 92), eligibility for disability pension is unobserved by the econometrician. Special method is developed to represent the unobservable doctoral screening process which has to recognize the inability of an individual to continue working due to health conditions and is the basis of the disability retirement. Thus, health becomes one of the most important notions and its modeling (described below) appears as the central innovation in the current paper.

The existing literature on retirement is generally in consensus about the important role of health in the retirement process. Many of the retirement studies which have established a strong influence of the economic incentives to the retirement decision neglect the health effects, although do acknowledge the need to take them into account (Gordon and Blinder, 1980; Blau, 1994; Hernæs et al., 2000; Borsch-Supan and Schmidt, 2000; Hernæs and Strom, 2000; Powers et al., 2001; Krueger et al., 2002; Chan et al., 2003; Conti et al., 2006). (Bloom et al., 2004) shows in a theoretical model how dropping out of labour market may be induced by worsening health. (Bound, 1998) takes this approach to the data and finds empirical evidence of health effects which are especially strong when measured on the relative rather than absolute scale. (Bound et al., 1995, 1996) show that controlling for health in a static model accounts for most of the racial gap in the labour force attachment and disability status of older American men and women in the 1990s. (Disney et al., 2006) use British household panel survey (1991-1998) to find that both current and lagged health shocks are positively correlated with a decision to drop out of labour force. (Au et al., 2005) also find evidence of the influence of health on the retirement behavior when they use Canadian data. After (Henretta, 1983; Hurd, 1989; Bourguignon and Chiappori, 1992a; Bourguignon and Chiappori, 1992b; Blau, 1997) showed the influence of a spouse to the individual retirement decision and the household approach was widely taken into the retirement research[b], (Coile, 2004) examines the "added worker effect" which suggests an increase of labour force participation of an individual after a negative health shock to the spouse. (Olson, 1998) studies American households and shows that wives without spousal health benefits are more likely to work full-time than those who are covered by the spousal health insurance. Health insurance is also shown by (Rogowski and Karoly, 2000) to be important for the retirement decisions, in particular, access to post-retirement health insurance has a large effect on retirement[c].

Once the significance of the health status on retirement behavior has been recognized, the issue of establishing a plausible and practical model that is capable of providing reliable simulations of the key policy measures becomes a major concern. (Gruber, 2000) studies the elasticity of labour force participation with respect to disability benefit generosity using a difference-in-difference approach and finds sizable labour supply response to possible changes in disability benefits. Studies based on the reduced from models, however, can not be applied to simulate the effects of central policy reforms. The goal of efficient policy development is best achieved with a structural approach to modeling labour market transitions which captures the existing state of nature not simply by establishing certain relationships among observed values, but instead through estimating more substantial stationary parameters of the processes driving the observed behavior. Therefore structural models are able to represent the responses of the labour market to a given policy and have been widely used in the retirement studies within different frameworks: static discrete choice (Dagsvik and Strøm, 1997; Dagsvik, 2002; Bratberg et al., 2004; Dagsvik and Strøm, 2006; Dagsvik and Jia, 2006), quasi-dynamic approach (Jia, 2000; Hernæs and Strom, 2001; Iskhakov, 2003), lifecycle approach (French, 1999; Gustman et al., 2004a, 2004b; Gustman and Steinmeier, 2005) and other (Hurd, 1989; Blau, 1997; Michaud and Vermeulen, 2004).

---

[b] See, for example, Gustman and Steinmeier (2000), Vermeulen (2002), Coile (2003), Jia (2003), Gustman and Steinmeier (2004), Hernæs, Jia and Strøm (2006).

[c] For the survey of the literature on the effects of health insurance see Gruber and Madrian (2002).

However, the occurrences of structural modeling of disability are quite rare in the existing literature. This is mainly because of the mentioned difficulty of modeling eligibility for disability pension which could in principle be overcome if reliable and transparent measures of health were available. Unfortunately, absence of essential data makes it impossible to incorporate the process of health screening into the structural model. Moreover, not only health data is hardly reliable and seldom available, but it also bears internal inconsistencies which are given a lot of attention in the literature.

Health can be measured in many different ways. The first and the most straightforward way is to ask respondents health-related questions in a survey and to use the answers for constructing either multidimensional or scalar measures usually on a simple ordinal scale (Bound et al., 1996; Dwyer and Mitchell, 1999; Kreider and Riphahn, 2000; McGarry, 2002; Heiss et al., 2003). These measures may suffer from multidimensionality and incomparability, scale simplicity (as pointed out by (Allison and Foster, 2004)) and other problems, but most of all they may suffer from endogeneity concerns described, for example, in (Bound, 1991). Together with other authors they raise the suspicion that answering questions about health status the respondents, especially those unemployed, may be rationalizing their labour market state or work preferences, which leads to overestimation of the influence of health in comparison with economic factors (known as "justification bias"). This suspicion is to some extend neutralized when the survey questions are less direct and address simple activities of daily living (ADL)[d] – health measures based on such questions are considered more "objective" (Heiss et al., 2003; Coile, 2004). Another possibility to eliminate the justification bias comes from introducing additional explanatory variables to instrument health. As pointed out by (Bound, 1991) in some circumstances, this gives even worse results, and therefore health indexes incorporating both subjective and objective information (possibly in different proportions) together with some individual characteristics may serve as "best" health measure (Bound, 1998; Dwyer and Mitchell, 1999). (Disney et al., 2006) call a version of such index a "health stock" and use it as one of the explanatory variables in a bigger labour market behavior model. The most serious drawback of a health index approach is its failure to provide a general theory of index construction – health indexes are task and project specific. Further, (Kreider, 1999) uncovers deeply imbedded inconsistency of health measurement which follows from the fact that health in medical sense is very different from work limitation measures important for labour market studies. The two concepts are very much related but when measurement is not perfect, the implied errors may be large. This is especially vivid when the health indexes are aimed at specific applications. Thus, for example, body mass index (BMI) or health utility index mark 3

---

[d] "Do you have problems walking up the stairs?" rather then "Is your health limiting your ability to work?"

(HUI3) used in the labour market analysis in (Rogowski and Karoly, 2000; Au et al., 2005) seem mostly medically oriented while (Benitez-Silva et al., 2004) give an excellent example of constructing a health index that outperforms the procedure used by Social Security Administration in the US for evaluating the disability applications. The last index is based on both the subjective measures (health related questions) and objective measures (ADL questions and individual characteristics) and is optimized for a narrow job. It is possible to come up with yet more "objective" measures for health as, for example, diagnosis (Gjesdal and Bratberg, 2003) or utilization of medical services (Currie et al., 1995). While the latter measure is definitely more appropriate for medical rather than labour market study, the former is used in a simplified form when health is measured by different health related events as stroke, cancer diagnosed, etc. (Heiss et al., 2003; Coile, 2004). Finally, mortality appears as the most objective but too aggregate measure of health and is therefore not used very often in microeconomic research (McGarry, 2002; Autor and Duggan, 2006). In the same time, self-assessed life expectancy is shown to be well-behaved measure (Hurd and McGarry, 2002) which also can be used in labour economic research (Coile, 2004).

Most of the different approaches for measuring health listed above originate in the corresponding survey questions. When instead it is a register data collected by the authorities which is used for estimation, the choice of health measures is very much reduced. The only available from the above options are mortality, medical services utilization and medical records. When neither of these is suitable or available, sick leave data from the employer registers may be the only option to serve as a health proxy in spite of unclear biases it may have in measuring work limitations (Gjesdal and Bratberg, 2003)[e].

The described controversies about the health measures are only magnified when it comes to modeling health dynamics. Increasing data requirements and multiplied measurement errors limit the available options and force the researches to simplify health related aspects of the models up to their complete elimination as in (Jia, 2005). (Bound, 1998; Au et al., 2005; Disney et al., 2006) reflect the dynamic aspects of health by using lagged health measures among the current period explanatory variables. (Gustman et al., 2002; Heyma, 2004) estimate complicated structural dynamic models but don't include health into the set of stationary variables and use health measures as exogenous. When health is allowed to vary over time, the movements may be very much restricted with the restrictions not necessarily implied by the theoretical setup but rather by the tractability considerations (Berkovec and Stern, 1991; Heiss et al., 2003). In those rare occasions

[e] In Gjesdal and Bratberg (2003) the number of days on sick leave is shown to be a significant predictor of the disability condition with the overall pseudo-R^2 in the model at about 30%.

when health is true endogenous state variable, it is usually measured on a simple dichotomous scale and is assumed to follow a Markovian motion rule (Rust and Phelan, 1997; French, 1999).

Bearing in mind the described drawbacks of different health measures and missing the essential health data of the necessary quality, I base the model on the new interpretation for health variable and use different approach to incorporate health into the structural model with disability retirement option. Let h denote the very eligibility for disability pension, so when $h = 1$ the option to take up disability pension appears in the individual choice set whereas when $h = 0$, there is no such option[f]. This definition completely eliminates the controversy of distinguishing medical aspects of health from labour market effects and allows me to assess the hidden medical screening process in the model. The variables although becomes unobservable in the data and is therefore treated as latent, in other words kept as a parameter while developing the model and integrated out on the estimation stage. Thus, health (equivalent to eligibility for disability) variable simply accounts for implied unobserved heterogeneity among the decision makers about their choice sets. State variable h is assumed to follow a simple Markov process which parameters are estimated within the general incomplete information maximum likelihood estimation procedure[g]. This approach makes the model numerically more complicated but still tractable and yields sensitive results.

In other respects the developed model estimated using the nested fixed point algorithm by (Rust, 1994) follows the tradition of structural stochastic dynamic programming originating in (Rust, 1987) and broadened for labour market analysis in (Rust, 1990; Rust and Phelan, 1997). The individuals are assumed to rationally maximize their expected discounted lifetime utility choosing at each period the best response to the evolving stochastic environment surrounding them. This environment is represented by the state vector sufficient to define at each period an individual choice set and feasible utility level dependent on the chosen alternative. Besides health, the state vector contains records of previous labour market state, current period job match, individual eligibility for the early retirement program, existence of a spouse indicator and two income proxies representing short and long term trends in the individual earnings. Preferences are represented with an indirect utility function which also counts for some heterogeneity contained in the state vector. Altogether the model contains 31 structural parameter some of which are related to the transition probabilities of the state variables and some of which enter the specification of the utility function.

---

[f] Later in the paper the health variable is defined more accurately and slightly differently.

[g] I could have used the EM algorithm (developed by Dempster, Laird and Rubin (1977) and reviewed in Ruud (1991)), but simple distributional specification of the latent processes facilitates estimation in one step using overall likelihood.

Once the structural parameters of the model have been estimated, it is straightforward to utilize them in policy simulations. Capability for revealing behavioral responses to a given policy not only in the period when it is implemented (as in static models) but also in other periods before and after the implementation constitutes the irrefutable advantage of the current dynamic model. In this paper I illustrate the application by investigation of the extent to which the AFP and disability exits can be viewed as substitutes. Potential substitution effect between the two retirement possibilities could to the great extent alter the effects of any particular policy aimed on either of them. Early retirement rules in Norway leave quite a lot of room for such interaction – as reported by (NOU 2004:1, 2004) about two thirds of the labour force has an option to take up AFP option up to five years earlier than usual retirement. This questions was addressed before in two separate papers and reflected some controversy. (Bratberg et al., 2004) adopt a non-parametric comparison along with a discrete choice model to investigate the question of interdependence of AFP and disability retirement and find clear signs for substitution effect of the magnitude between 8.6% and 22.4%. At the same time (Røed and Haugen, 2003) using a quasi-natural experiment of lowering the early retirement age find practically no substitution between the two exit routes which is line with a previous study on American data by (Bound, 1989). Neither of the two papers assessed underlying changes in the health status and concealed eligibility for disability pension – the factors that come into play exactly when one of the retirement options becomes unavailable. Therefore the current paper allows me to shed some more light of this issue and trace the dynamic consequences of elimination of the early retirement.

The rest of the paper is organized as follows. The second section describes the theoretical model and the estimation technique, section three is devoted to the data description, section four – to the final empirical specifications for the model. Sections five and six present calibration and estimation results respectively. Last section presents policy simulation and is followed by the concluding remarks.

2. *The model*

The paper develops a discrete time structural dynamic programming model built on the assumption that individuals maximize expected discounted lifetime utility in order to find the optimal path of transition from work to retirement. Health status is modeled as underlying latent stochastic Markov process which alters the set of alternatives available to decision makers at each period. This section starts with formulating individual agent problem and ends with the expressions for the choice probabilities and the likelihood function.

2.1. The agent problem

The main purpose of the model I develop is to represent the process of making the retirement decision and to answer the question when old-age individuals withdraw from the labour force. I start with defining a single agent decision rule which subsequently leads to the choice probabilities.

Let vector $s_t \in S$ contain the values of the *state variables* corresponding to the full set of social and economic factors effecting the agent's decision making at period *t* where S is the corresponding state space. Some of these variables evolve over time in a random fashion forming a stochastic process $\{\tilde{s}_t\}$ which can be at least partially controlled by a *decision variable* d_t, representing agents actions in response to the unrolling realization of the state process $\{\tilde{s}_t\}$. Assume that the agent acts rationally trying to maximize a time separable discounted objective function

$$\tilde{U} = \sum_{t=T_0}^{\tilde{T}} \beta^{t-T_0} U(\tilde{d}_t, \tilde{s}_t), \tag{1}$$

where $U(d_t, s_t)$ is an instantaneous indirect utility at period t and β is an intertemporal utility discount factor. The tildes emphasize the fact that corresponding variables are stochastic: the decision variable is a function of the uncertain state history $\tilde{d}_t = \delta_t(\tilde{s}_t, \tilde{s}_{t-1}, \tilde{s}_{t-2}, ..)$, that is the best response to the current and possibly previous states. A set of these functions forms a *decision rule* $\delta = \left(\delta_{T_0}, ..., \delta_T\right)$ describing agents decision making at each time period and thus inducing stochastic process $\left\{\tilde{d}_t, \tilde{s}_t\right\}_\delta$ which starts in the initial point (d_{T_0-1}, s_{T_0-1}).

For convenience the time index in the model serves as indicator of age, thus random $\tilde{T}$ in (1) indicates the age of death. In the dynamic programming specification which follows it is substituted in a standard way with the survival probabilities, and for the empirical implementation of the model I fix the limits T_0 and T such that the most important life span for retirement behavior study is covered: $T_0 - 1 = 50$ to include sufficient number of years before possible retirement in order to capture planning and health dynamics and $T = 70$ due to the compulsory retirement age after which no transfers occur and no decisions are made. Other dynamic programming studies include later ages into the model to make it more realistic in a sense that each individual is tracked up until certain death – then the time horizon is set so that all agents surely die within the modeling period. The associated calculation burden can be escaped by limiting the modeled period with the necessity to estimate additional termination function $\Lambda(\tilde{s}_T)$ which captures the remaining lifetime utility. Thus the latter approach does have its computational downside as additional parameters are introduced, but this complication seems to be of the smaller scale compared to setting a large time horizon, and therefore is adopted in the paper.

An important assumption that has to be made concerns agents preferences over uncertainty. I assume the decision maker to be an expected utility maximizer. This assumption is in a sense inevitable because as (Rust, 2006) discusses the expected utility concept is quite deeply imbedded into the dynamic programming methodology itself. I assume that when taking decisions regarding their labour market state individuals solve the following sequential decision problem:

$$E\left\{\sum_{t=T_0}^{T}\left(\prod_{\tau=T_0}^{t}\rho_\tau\right)\beta^{t-T_0}U(d_t,s_t)+\Lambda(s_T)\right\} \underset{\delta\in\mathfrak{F}}{\to} \max \text{ [h]}, \tag{2}$$

where expectation is taken over the survival probabilities ρ_τ (corresponding to survival from period $\tau-1$ to period τ) and the set of subjective transition probabilities $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ that govern the stochastic process $\{\tilde{d}_t, \tilde{s}_t\}_\delta$ induced by the given decision rule $\delta \in \mathfrak{F}$. I deliberately restrict the model to express Markovian property by limiting the influence of history of states in the subjective transition probabilities – this common simplifying assumption allows for drastic reduction of the computational burden on the estimation stage. In the same time this approach has been considered plausible in socio-economic studies as there are reasons to believe that human behavior is conditioned on the current life situation to much greater extent compared to the events in the past. Along with the assumption that decisions are made within each period after the realization of the state variables, the state history in the expression for decision rules can be dropped apart from the first component denoting the current state. Note also that some of the transition probabilities may be degenerate if a state variable evaluates according to some deterministic low of motion.

Agent problem (2) restricts the choice of decision rule in the maximization procedure to class $\mathfrak{F}$ which in the case of Markov decision problem in the finite time horizon can be limited according to

[h] Notation introduced by Nobel prize winner L.V. Kantorovich (see, for example, Kantorovich (1976)).

the theorem 2.1 in (Rust, 1994) to the class of feasible Markovian decision rules of the form $\delta = \left(\delta_{T_0}(\tilde{s}_{T_0}),...,\delta_T(\tilde{s}_T)\right)$. Under the weak regularity conditions the theorem guarantees the existence of Markovian non-randomized optimal decision rule δ^* that solves the agent problem (2).

Even though the optimal decision rule δ^* is a deterministic function of the state, the agent is not pre-committed to any set of fixed actions designed ax ante (as he would be in a life cycle model), instead in each period the agent makes an optimal decision using the new information emerging over time. Still, the commitment to the optimal decision rule requires the agent to act in a time-consistent fashion, in other words, all the decisions taken in the past are perceived by the agent as optimal and thus there are no incentives to change the part behavior ex post. This coincides with the assumption of expected utility maximization, in fact (Hammond, 1988) shows that under the time consistency assumption (along with some technicalities) expected utility concept is the only feasible representation of preferences.

Feasibility conditions that define the class $\mathfrak{F}$ are expressed in a family of choice sets $D_t(s_t,d_{t-1})$ that represent the available to the agent options at period *t*. Decision rule $\delta = \left(\delta_{T_0},...,\delta_T\right)$ is said to be feasible if and only if for each $t \in \{T_0,...,T\}$ $\delta_t(s_t) \in D_t(s_t,d_{t-1})$. In other words, the class $\mathfrak{F}$ can be represented by a Cartesian product of the choice sets $\mathfrak{F} = \otimes_{t=T_0}^{T} D_t(s_t,d_{t-1})$. Definitions of the family of choice sets $\{D_t(s_t,d_{t-1})\}_{t\in\{T_0,..,T\}}$ and transition probabilities $\left\{p(s_t \mid s_{t-1},d_{t-1})\right\}$ conclude the agent sequential decision problem setup.

The family of the choice sets will be defined shortly, while the set of transition probabilities is fully defined only in the third section of the paper since it is deeply connected to the empirical specification of the model. For some state variables however the evolution rules are simple and can be presented sooner apart from the data.

2.2. Solution technique for the agent problem

The sequential decision problem (2) falls into the mathematical category of stochastic optimal control in discrete time. The problems of this type rarely have analytical solutions. Instead a much more common technique may be applied, namely the numerical method of calculating optimal decision rule through backward induction. In order to proceed, define a value function $V_t(s_t)$ by

$$V_t(s_t) = \begin{cases} \max\limits_{d_T \in D_T(s_T,d_{T-1})} \left[U(d_T,s_T) + \Lambda\left(s_T\right)\right], t = T, \\ \max\limits_{d_t \in D_t(s_t,d_{t-1})} \left[U(d_t,s_t) + \rho_{t+1}\beta \sum\limits_{s_{t+1}\in S} V_{t+1}(s_{t+1})p(s_{t+1} \mid s_t,d_t)\right], t < T, \end{cases} \tag{3}$$

and derive recursive expression for the optimal decision rule $\delta^* = \left(\delta^*_{T_0},...,\delta^*_T\right)$

$$\delta^*_t(s_t) = \begin{cases} \underset{d_T \in D_T(s_T,d_{T-1})}{\arg\max} \left[U(d_T,s_T) + \Lambda\left(s_T\right)\right], t = T, \\ \underset{d_t \in D_t(s_t,d_{t-1})}{\arg\max} \left[U(d_t,s_t) + \rho_{t+1}\beta \sum_{s_{t+1}\in S} V_{t+1}(s_{t+1}) p(s_{t+1} \mid s_t, d_t)\right], t < T. \end{cases} \tag{4}$$

Expressions (3) and (4) allow for the computation of the optimal decision rule explicitly starting from the last period (first cases in the expressions) and continuing backwards step by step using already calculated values of the value function in the next consecutive period. Proceeding this way ensures that the *optimality principal* that characterizes solutions of the optimal control problems is satisfied. The optimality principle states that at each step the optimal control solves the corresponding sub-problem which starts at the current step and reproduces the original one up to the period *T*. In other words, the optimality principle states that decision rule $\delta^* = \left(\delta^*_{T_0},...,\delta^*_T\right)$ is optimal if for every $\tau \in \{T_0,..,T\}$

$$\left(\delta^*_\tau,...,\delta^*_T\right) = \underset{(\delta_\tau,..,\delta_T)\in\otimes_{s=\tau}^T D_s(s_s,d_{s-1})}{\arg\max} E\left\{\sum_{t=\tau}^T \left(\prod_{s=T_0}^t \rho_s\right)\beta^{t-\tau} U\left(\delta_t(\tilde{s}_t),\tilde{s}_t\right) + \Lambda\left(\tilde{s}_T\right)\right\}. \tag{5}$$

It is straightforward to verify that a backward induction algorithm indeed produces the sequence of optimal controls that altogether constitute a numerical representation of the optimal decision rule for a particular realization of the $\{\tilde{s}_t\}$ process corresponding to a given agent. It is also obvious that the value function $V_t(s_t)$ takes the values of the optimized objective function in each of the sub-problems given by (5), and when $\tau = T_0$

$$V(s_{T_0}) = \max_{(\delta_{T_0},..,\delta_T)\in\mathfrak{F}} E\left\{\sum_{t=T_0}^T \left(\prod_{s=T_0}^t \rho_s\right)\beta^{t-T_0} U\left(\delta_t(\tilde{s}_t),\tilde{s}_t\right) + \Lambda\left(\tilde{s}_T\right)\right\}, \tag{6}$$

$$\delta^* = \underset{(\delta_{T_0},..,\delta_T)\in\mathfrak{F}}{\arg\max} E\left\{\sum_{t=T_0}^T \left(\prod_{s=T_0}^t \rho_s\right)\beta^{t-T_0} U\left(\delta_t(\tilde{s}_t),\tilde{s}_t\right) + \Lambda\left(\tilde{s}_T\right)\right\}. \tag{7}$$

Thus, the optimal decision rule δ^* indeed constitutes the solution to the agent sequential decision problem.

2.3. Decision and state variables

The next several sections describe the essential parts of the agent sequential decision problem in greater details. First, the state variable vector s_t is populated with meaningful content and then the family of choice sets $\{D_t(s_t,d_{t-1})\}_{t\in\{T_0,..,T\}}$ and the transition probabilities $\left\{p(s_t \mid s_{t-1},d_{t-1})\right\}$.

Constructing the set of state and decision variables is not a simple process and is heavily effected by the following two considerations. First, the model must be adequate in describing the real empirical processes which are the subject of the study. This is crucial for getting a reasonable goodness of fit and respectively gaining the explaining and forecasting power of the model. Policy simulations which is one of the main objectives in the work, would be impossible with poor correspondence between the model and the reality. These considerations drive the desire to make the model design as close to the reality as possible, taking into account many personal characteristics of the agents and those of the states on the labour market. On the other hand the model must be realistic in computational sense. Well studied problem of exponential grows of the amount of calculations required to solve the agents problem with the dimensions of the state and decision space named by (Bellman, 1962) "the curse of dimensionality" prevents use of too realistic setups even with the drastic development of computing technology after the curse was first encountered. In the current paper I adopt a compromise which is based on the fact that bounded discrete variables add much less to the dimensions of the problem (when it is expressed in the total number of grid points of the state space) than the continuous ones. Indeed, any continuous state variable must be represented by a grid vector (random or regular) which most likely contains much more points than any bounded discrete variable represented by finite number of its values. This is specially true when the discrete variables are defined with small number of values. Under this logic I mostly use discrete variables to represent different systematic and stochastic aspects of an individual working life between ages 50 and 70 which were mentioned in the introduction.

The constructed set of state variables naturally separates into several categories. The age of the agent is worth mentioning in the first category. This is an essential variable for most economic processes cointegrated with labour market transitions, but is omitted from the state variables vector because as mentioned above it is identical to time index. This became possible by fixing specific age window so that at the first period in the model corresponds to 50 years of age.

The second category is the subset of state variables that could by convention be called constants. These are the agent specific characteristics which are constant throughout the modeling period but have to be kept in mind when calculating the likelihood function. The most important (and the only two used in the empirical specification of the model) are individual specific AFP age that is the earliest age of possible retirement through the early retirement program[i] and gender of the decision maker. To simplify the notation I omit these variables for the rest of the paper except for after the empirical part, but it should be kept in mind that the agents are differentiated in accordance to them.

---

[i] This variable would also serve as an indicator for the population cohort because birth year of the agent is not controlled for and timeline is synchronized with age.

The third category is defined as the variables effecting the choice set in the current period. These are previous labour market state, health, job match and eligibility for early retirement. The fourth and final category contains the variables that together with current decision determine the economic situation in the current period and thus effect current period utility. These variables are spouse existence, number of last consecutive years with high incomes and aggregate wage[j]. Before describing the variables from the last two categories in details consider the timing assumption adopted in the paper.

I assume that the state process $\{\tilde{s}_t\}$ is evaluated in the beginning of each period followed by the reaction on the decision maker's side, so that the values of the current period state variables are realized before the decision is chosen. Decision made afterwards is only capable of influencing the utility level for the current period and alter somewhat the evaluation of the state process in the next period. Note that this very assumption was implied above in the sequential decision problem description. The purpose of this setup is to emphasize that the state process is in a sense underlying and superior to the agents behavior. The agent is only forced to react to the changes in his or her current situation trying to obtain the highest possible level of utility from the realized conditions. This seems to be a more reasonable description of the average later working life than the opposite one. Aged workers are more likely to be pursuing their retirement plans suffering from sudden labour market moves or health problems rather than pursuing their carrier ambitions. Besides, the state variables describing the economic situation of an individual by age 50 have already gained certain momentum making them harder to control, for example the underlying aggregate wage profile already at the initial period contains most of the working history of the individual and can only be corrected other than shaped from scratch.

Consider first the variables constituting the choice set in the current period. These are:

- Previous period labour market state $ps_t \in \{0,..,7\}$ is the main choice set defining variable that indicates what labour market states are available in the current period under the important absorbing assumptions explained shortly. To facilitate the chosen timing structure it is essential to include lagged labour market state variable into the state vector. Note however that this trick does not compromise the Markovian structure of the decision problem. For the purpose of the current study the following labour market states are introduced[k]:
    - $ps_t = 0$ – out of labour market (OLM),

[j] Exact definition for the aggregate wage is given in the later sections.

[k] The chosen labour market states, and specifically combination of work and disability, are suggested by the preliminary data analysis.

- $ps_t = 1$ – full time early or regular pension,
- $ps_t = 2$ – full time disability,
- $ps_t = 3$ – unemployment (including partial unemployment),
- $ps_t = 4$ – employment in non-AFP company[l],
- $ps_t = 5$ – partial employment in non-AFP company, partial disability,
- $ps_t = 6$ – employment in AFP company,
- $ps_t = 7$ – partial employment in AFP company, partial disability.

The first three labor market states[m] constitute the out of labour market group while the last five correspond to active labour market positions. In order to keep track of availability of early retirement option I distinguish AFP from non-AFP labour market participation. Partial disability is the only considered form of partial retirement, other types of phased retirement are assumed away. OLM state allows building exhaustive set of labour market states which is required in the discrete choice models of many sorts. Economic interpretation of this labour market state is not working at all (house wife) or self-employment. Mutual independence of the introduced labour market states as well as strict definitions for them are given in section 3 of the paper, while the absorption assumptions are explained below within the definitions of choice sets.

- Health status $h_t \in \{0,1,2\}$ is the first and the main latent variables in the model. As it was described in the introduction existing literature rarely addressed health directly because this very important state variable is extremely hard to measure. Given considerations drove me to introduce health as a latent variable and give it a special definition. Health is thought of specifically as eligibility for disability pension. This way, $h_t = 0$ is good health with no option to retire through disability, $h_t = 1$ gives an option to become partially disabled with reduced labour market opportunities while $h_t = 2$ implies full time disability. Thus, some of the health status should be recoverable from the data on occupied labour market states, but the hypothesis that bad health could be in some sense concealed until a convenient retirement opportunity comes around makes the health variable unrecoverable completely from the data

[l] AFP and non-AFP companies differ in their participation in the early retirement agreement, see Appendix A.1, p. 92.

[m] I decided to distinguish "state" in the dynamic programming sense from "labour market state" by keeping the latter as definitive expression even though some labour market states actually indicate the absence from the labour market.

and gives rise to the question of possible substitution of disability with other forms of retirement.

- The need to explain transitions between AFP and non-AFP employment when the former for any preferences specification dominates the latter (because of the retained option to retire early) as well as the need to explain transitions to unemployment facilitates the use of separate labour market matching process $m_t \in \{0,1,2\}$ with the following interpretation. If $m_t > 0$ there is a job opening in the current period ($m_t = 1$ – in non-AFP, $m_t = 2$ – in AFP company), otherwise an individual is forced to unemployment and possibly to full time disability. This is the second latent variable which is however fully recoverable from the data on occupied labour market states.
- Adequate representation of the complicated early retirement process in Norway (see section A.1 in the Appendix, p. 92) requires a special variable $e_t \in \{0,1\}$ to keep track of the eligibility conditions not directly verifiable within the model. Since affiliation to a AFP company is directly verifiable by the labour market state, this variable mainly reflects individual AFP eligibility criteria. Thus, $e_t = 1$ is necessary but not sufficient for the early retirement option to be included in the choice set in the current period. More detailed description of how the verification of the AFP rules is performed within the state vector is given in section 4 when the motion rules of variable e_t are defined.

The way these four state variables determine the current period choice set is described in the next section after the definitions of state and control variables is completed. The final category of state variables contains variables essential for current period utility calculation which are:

- Spouse existence indicator $sp_t \in \{0,1\}$: if $sp_t = 1$ the agent under consideration is not a single individual but a household which by the simplified construction of the model differs from a singe person household (individual) only by possible existence of additional income source from a spouse. I assume that full households are governed by the same preferences as single households which is justified by the unitary or Stackelberg equilibrium approaches to household preference modeling (Hiedemann, 1998; Jia, 2003; Hernæs et al., 2006). I also assume away the events of new marriages but allow full households to become single in case of divorce or death of the spouse (which is more probable is the considered age group). Further discussion is given in section 3 dealing with data issues.
- Finally, the aggregate wage $aw_t \in \mathbb{R}_+$ represents the lifetime trend in the wage income flow for the agent. This is the only continuous state variable in the model and bears most of the

burden of explaining household income in a particular time period. Section 3 considers several candidates for this variable available in the data and discusses related issues.

- In addition, bearing in mind that the rules for social benefits (pensions, disability) include conditions on the number of consequent years with high income (where "high" is defined by the level of basic pension G) I decided to include in the state vector additional discrete variable indicating the number of last consecutive years with wage over the basic pension amount $nw_t \in \{0,1...10\}$. This number is truncated at 10 in accordance with the mentioned principle of keeping the problem dimensions possibly lower and under assumption that additional values for this variable do not bear considerable additional information. This variable will however be able to improve the information fullness of the state vector indicating the short term trend in the household income flow.

This concludes the description of the state variables defining the state vector s_t as

$$s_t = (ps_t, h_t, m_t, e_t, sp_t, nw_t, aw_t). \tag{8}$$

In contrast to the state vector, the decision variable in the current model is unitary. Define the set of all possible decisions (decision space) $D = \{0,..,4\}$ with the following interpretation.

- $d_t = 0$ – the agent remains on the labour market, does not apply for any pension,
- $d_t = 1$ – the agent applies for disability benefits, but remains on the labour market,
- $d_t = 2$ – the agent retires, applies for disability benefits,
- $d_t = 3$ – the agent retires, applies for old age or AFP pension,
- $d_t = 4$ – the agent leaves labour market, but does not apply for any pension.

Thus, the decision variable indicates the intentions of the agent to acquire a certain position on the labour market, which is matched against current state to determine actual outcome (which becomes the current labour market state and is recorded and next period ps_{t+1} variable). Decision is made in two dimensions: whether to stay on the labour market or leave it, and whether to apply for pension or disability benefit. One combination of the answers (namely, staying on the labour market and simultaneously receiving pension or phased retirement) is ruled out, which leaves 5 possible values for the decision variable. Clearly, some intentions are also useless in particular situations, for example intention to go on pension before the early retirement age, such cases are regulated in the model by the definitions of the choice sets discussed in section 3.

According to the introduced categories of the state variables the expression $D_t(s_t, d_{t-1})$ may be concretized as $D_t(s_t, d_{t-1}) = D_t(ps_t, h_t, m_t, e_t, d_{t-1})$ and further as

$$D_t(s_t, d_{t-1}) = D_t(ps_t, h_t, m_t, e_t), \qquad (9)$$

because the influence of the previous period decision d_{t-1} to the choice set is recorded in the current period state variable ps_t. By construction, $D_t \subset D$ for any $t \in \{T_0,..,T\}$.

2.4. Decision tree

Definitions of the choice sets from the family $\{D_t(ps_t, h_t, m_t, e_t)\}_{t\in\{T_0,..,T\}}$ are very much related to the motion rules of the ps_t state variable. Given the current period state vector s_t, a feasible decision $d_t \in D_t$ is chosen to determine the current period labour market state that is recorder in the ps_{t+1} variable. Thus, defining the family $\{D_t(ps_t, h_t, m_t, e_t)\}_{t\in\{T_0,..,T\}}$ is equivalent to setting restrictions on the evolution of labour market state. The following considerations then shape the choice sets.

- Mandatory retirement age in Norway is 70, hence only retirement is available among all the labour market states at this age.
- Before the usual retirement age of 67 the only possible form of retirement is early retirement (AFP), after the usual retirement age of 67 any individual can go on pension (in particular, full time disabled are forced into pension).
- Early retirement through AFP program is only available for the eligible individuals. As described in the Appendix (p. 92), both company and individual criteria have to be met. Employment in the AFP-affiliated company and some other directly verifiable requirements can be checked with the means of the state variables (first of all ps_t and nw_t) while non-directly verifiable requirements as individual eligibility is tracked with state variable e_t.
- Full time and partial disability can only be attained when the health variable h_t takes correspondingly values of 2 and 1. This obvious feature of choice set expansion with bad health is the central feature of the model. Note however, that the matching process $\{\tilde{m}_t\}$ does not modify the choice sets, instead it separates intentions of an individual from the actual labour market state attained in the outcome[n]. Also “very bad” health $h_t = 2$ contracts the choice set to a singe point (up to retirement age) $d_t = 2$ instead of expanding it.

Finally, the following absorption assumptions are made.

- Pension is completely absorbing, once a person is retired, he or she neither may go back to work nor has incentive to transfer to any other state other then pension.

[n] Except for the situation of bad health and no job match – then the individual is forced into the full time disability.

- Full time disability is equivalent to completely absorbing, as the only transfer from this state is inevitable transfer to pension at age 67.
- Once in OLM state an individual is not allowed to return to the labour market neither as an employee nor as a registered unemployed[o].

Absorption assumptions are rather strict – they completely separate the active area among the labour market states so that once an agents takes a decision to leave active labour force (states on the labour market from 3 and up), no return is possible. I impose these restrictions to emphasize the directional nature of the retirement decision. Independent of the chosen form of retirement this very decision significantly reduces the choice set of an agent for the rest of his or her life. For elderly workers over 60 this is completely substantiated in the data, in the younger group such trend is well pronounced. These assumptions have also been used in the previous studies, for instance (Stock and Wise, 1990b; Stock and Wise, 1990a; Hernæs and Strom, 2001).

In order to force these restrictions and form the decision tree, it is necessary to introduce the motion rule for the labour market state variable. It is best described in a table (see Table 1). Here the correspondence between decision variable d_t and the resulting labour market state in the current period ps_{t+1} is separated with a "filter" of state variables representing all the conditions described above.

Table 1. Evaluation of current labour market state[p].

Control			Filter					Resulting labour market state ps_{t+1}	
d_t	Remain on LM?	Apply for pension?	ps_t	h_t	m_t	e_t	Age		
0	yes	no	≥ 3	0	0	-	< 70	3	Unemployment
				$\neq 2$	1			4	Non-AFP employment
					2			6	AFP employment
1	yes	disability	$\neq 1$	1	0	-	< 70	2	Full time disability
			≥ 3	1	1			5	Partial disability (non-AFP)
				1	2			7	Partial disability (AFP)
2	no	disability	$\neq 1$	> 0			< 70	2	Full time disability
3	no	AFP/NIS	≥ 6	-	-	1	$\geq afp$	1	Pension
			$= 1$			-	$\geq afp$		
			-				≥ 67		
4	no	no	$\neq 1, 2$	$\neq 2$	-	-	< 70	0	OLM

[o] This limitation is suggested by the preliminary analysis of the data and allows for considerable simplification of the likelihood function and the backward induction calculation.

[p] In the forth row the age is compared against the individual early retirement age.

Intuitive definitions of the choice sets are then very simple. A given value of d_t is included in the choice set if and only if it passes through the "filter", or in other words leads to a certain current period labour market state when the state vector is fixed at its current value. Expressions (10-14) give strict definition of the family $\{D_t(ps_t,h_t,m_t,e_t)\}_{t\in\{T_0,..,T\}}$.

$$d_t = 0 \in D_t(ps_t,h_t,m_t,e_t) \Leftrightarrow \begin{cases} ps_t \geq 3, \\ h_t \neq 2, \\ t < 70, \end{cases} \tag{10}$$

$$d_t = 1 \in D_t(ps_t,h_t,m_t,e_t) \Leftrightarrow \begin{cases} h_t = 1, \\ t < 70, \\ ps_t \neq 1, \text{if } m_t = 0, \\ ps_t \geq 3, \text{if } m_t \in \{1,2\}, \end{cases} \tag{11}$$

$$d_t = 2 \in D_t(h_t,m_t,e_t,s_{t-1}) \Leftrightarrow \begin{cases} ps_t \neq 1, \\ h_t \in \{1,2\}, \\ t < 70, \end{cases} \tag{12}$$

$$d_t = 3 \in D_t(h_t,m_t,e_t,s_{t-1}) \Leftrightarrow \begin{cases} ps_t \geq 6, \\ e_t = 1, \quad or \begin{cases} ps_t = 1, \\ t \geq afp, \end{cases} or\ t \geq 67, \\ t \geq afp, \end{cases} \tag{13}$$

$$d_t = 4 \in D_t(h_t,m_t,e_t,s_{t-1}) \Leftrightarrow \begin{cases} ps_t \in \{0,3,4,5,6,7\}, \\ h_t \neq 2, \\ t < 70, \end{cases} \tag{14}$$

where *afp* marks the individual AFP retirement age.

It is easy to see that the family $\{D_t(ps_t,h_t,m_t,e_t)\}_{t\in\{T_0,..,T\}}$ is defined in such a way that for some combinations of state variables a corresponding choice set is empty. This may present a considerable problem for the calculation procedure implementing the backward induction algorithm which finds an optimal response for each combination of state variables at each period. The problem can be solved by assigning zero probabilities to the useless combinations of state variables. How this is done in the model and what advantages follow from this complication is described in the next section.

2.5. Motion rules

In this section I start defining the set of transitional probabilities $\{p(s_t \mid s_{t-1},d_{t-1})\}$ with the discussion of general assumptions about its structure and present simple motion rules. Definition is completed when the empirical specification of the model is given later in the paper due to the fact that many of the transition probabilities will be defined through statistical models estimated on the available data.

Generally speaking, transition probabilities $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ governing the Markov stochastic process $\{\tilde{d}_t, \tilde{s}_t\}_\delta$ constitute a square matrix (of the dimensions equal to the number of elements in the state space) which can be altered by the previous period decision d_{t-1}. This matrix is quite extensive. The defined value sets for the state variables result in the number of elements equal to $19\,008^2$ times the squared number of grid points for aw_t. This makes it practically impossible to estimate individual probabilities from the data. One solution to the problem suggested by (Rust, 1990) is decomposition of the full matrix to smaller blocks by assuming some specific dependence structure on the stochastic elements included into the model.

I make the following major assumptions with regard to the dependence structure. First of all, health is thought of as fundamental underlying Markov process, which evolves completely independently with the transition probability matrix

$$\{\pi_{ij}^{(h)}\}_{i,j\in\{0,1,2\}} = \begin{bmatrix} \pi_{00}^{h} & \pi_{01}^{h} & \pi_{02}^{h} \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}, \tag{15}$$

where unspecified matrix elements are parameters. Health transition probability matrix is uniform for all individuals. Bad health statuses are completely absorbing. This is very natural in light of health definition as eligibility for disability benefits, and also substantiated by the data[q].

I assume matching process to depend on the health in a way that would reflect limited labour market opportunities for the partial disabled. Consider transition probability matrix for $\{\tilde{m}_t\}$ of the form

$$\{\pi_{ij}^{(m)}\}_{i,j\in\{0,1,2\}} = \begin{bmatrix} \pi_{00}^{m} + (\pi_{01}^{m} + \pi_{02}^{m})\cdot(1-\theta^{bhm}) & \pi_{01}^{m}\cdot\theta^{bhm} & \pi_{02}^{m}\cdot\theta^{bhm} \\ \pi_{10}^{m} + (\pi_{11}^{m} + \pi_{12}^{m})\cdot(1-\theta^{bhm}) & \pi_{11}^{m}\cdot\theta^{bhm} & \pi_{12}^{m}\cdot\theta^{bhm} \\ \pi_{20}^{m} + (\pi_{21}^{m} + \pi_{22}^{m})\cdot(1-\theta^{bhm}) & \pi_{21}^{m}\cdot\theta^{bhm} & \pi_{22}^{m}\cdot\theta^{bhm} \end{bmatrix}, \begin{cases} \theta^{bhm} = 1 \text{ if } h_t = 0, \\ 0 \le \theta^{bhm} \le 1 \text{ if } h_t = 1, \\ \theta^{bhm} = 0 \text{ if } h_t = 2. \end{cases} \tag{16}$$

Parameter θ^{bhm} redistributes the probability mass away from the second two columns corresponding to the job openings independently from the previous value m_{t-1}. This insures that

---

[q] In fact, the transition probability from "bad" to "very bad" health π_{12}^{h} was assigned zero value in a preliminary calibration. This is due to the fact that both of these health statuses present a path to the same labour market outcome of full time disability (according to Table 1) and thus $\pi_{12}^{h} = 0$ does not compromise the flexibility of the model.

once agent's health is deteriorated the labour market opportunities can be immediately reduced. Note though, that the matching process transition probabilities are only dependent on health status. Other sources of heterogeneity (including individual heterogeneity) is not controlled for similar to the health status evolvement. When $h_t = 2$ variable m_t becomes zero with probability one.

The rest of the motion rules are less obvious. State variables e_t, nw_t and aw_t have more complicated dependence patterns which will be uncovered later in the paper. Spouse existence indicator sp_t is a Markov process with the fixed exogenous transition probabilities described in section 3. Evaluation of labour market state (which is recorded with lag one in the variable ps_{t+1}) is deterministic and has been described in the previous section with the means of Table 1.

As it was noted in the end of the last section some combinations of the state variables are useless in a sense that they are accompanied with empty choice sets and must be assigned zero probabilities in the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$. The other possible source of zeros in the matrix are deterministic or limited motions of the state variables, for example evolvement of health variable constrained by absorption (with zeros in (15)) or limited evolvement of nw_t which can in one period be either incremented by one or nullified (probabilities corresponding to the rest of the transitions are zeros). In general occurrence of zero elements and their distribution in the matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ should be carefully studied and may be used to speed up solution algorithm since skipping the calculations associated with the zero-probability states spares calculation time.

This is especially relevant if the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ contains complete rows and column of zero elements associated with improbable combinations of state variables. The following considerations should be utilized.

- The first time retirement possible in principle is at AFP retirement age. Therefore $\Pr\{ps_t = 1, t \le afp\} = 0$.
- Without loss of generality it is possible to enforce $\Pr\{e_t = 1, t < afp\} = 0$ and $\Pr\{e_t = 1, t \ge 67\} = 0$ because individual AFP eligibility indicator is essential in the model only between AFP eligibility age and normal retirement age.
- Moreover, $\Pr\{e_t = 0, ps_t = 1, t < 67\} = 0$ because AFP retirees are automatically AFP eligible. This is a limitation similar to the pension absorption, but expressed in a slightly different way.
- $\Pr\{e_t = 1, ps_t \notin \{1, 6, 7\}\} = 0$ because the AFP eligibility rules require AFP employment for the possible early pensioners in the last years prior to retirement which makes it impractical to

keep track of other AFP eligibility criteria (see Appendix for AFP rules). The only exception is absorbing retirement state itself.

- Without loss of generality $\Pr\{m_t \neq 0, ps_t \in \{0,1,2\}\} = 0$ since matching is only essential for the active labour market states and in the same time out of labour market states are jointly absorbing (if this was not the case, the initial match could be reset which would result in the loss of information).
- Similarly $\Pr\{m_t \neq 0, h_t = 2\} = 0$ because match is not relevant for fully disabled and this labour market state is absorbing.
- $\Pr\{h_t = 0, ps_t \in \{2,5,7\}\} = 0$ combines the logic from health transition probability matrix (15) and Table 1 to express explicitly that once on partial or full-time disability an individual can not happen to be in good health again.

Investigated improbable states constitute complete columns and rows of zero values in the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ which implies that these combinations of state variables can never be occupied by the modeled stochastic process $\{\tilde{d}_t, \tilde{s}_t\}_\delta$ and therefore can be completely eliminated from the calculation procedures.

Achieved reduction of state space is graphically shown in Figure 11 in the Appendix (p. 101). It displays the map of essential state variables, where the vertical axes corresponds to all possible values of the partial state vector (ps_t, h_t, m_t, e_t) and the horizontal axes represents age. Black areas are the only combinations with positive probabilities of occurrence. Exclusion of some state variables combinations allows for the reduction of the effective number of points for the value function calculation from 66 528 (times the number of grid points for aw_t) to 17 446, thus giving a 73,8% reduction in the dimensions of the problem!

The drawback of the state elimination technique is the need to very carefully implement the probability calculation procedure to ensure that eliminated states are indeed associated with zero probabilities. I will come back to Figure 11 and the complications related to it when the discussion of transition probabilities family $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ is continued later in the paper.

2.6. Likelihood function

To conclude the description of the model I follow (Rust, 1994) to present briefly the model estimation method.

To be able to rationalize inevitable discrepancies between the theoretical model and the observed data a standard approach is to suppose that an observed state vector is not complete, so that there are some additional state variables which effect the choices made by the agent but are not seen by the econometrician. In particular assume there exist five additional unobserved state variables combined in a vector $\varepsilon_t \in R^{|D|}$ which alter the instantaneous utility function $U(d_t, s_t)$ in the agent problem (2) so that

$$U(d_t, s_t) = u(d_t, s_t) + \varepsilon_t[d_t], \tag{17}$$

where $u(d_t, s_t)$ is non-stochastic component of the utility that will be given particular functional form later on and $\varepsilon_t[d_t]$ is the component of vector ε_t corresponding to the decision d_t. Important conditional independence assumption on the transition probabilities of the new state vector has to be made.

$$p(s_t, \varepsilon_t \mid s_{t-1}, \varepsilon_{t-1}, d_{t-1}) = q(\varepsilon_t \mid s_t) \cdot p(s_t \mid s_{t-1}, d_{t-1}). \tag{18}$$

Condition (18) extends dependence structure imposed above on the transition probabilities $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ to include the newly introduced unobserved variables. They are assumed to depend on their previous values only through the observed state variables and moreover constructed earlier transition probabilities for the observed state variables are completely independent of the newly introduced unobserved variables (although the decision may as well be dependent on them).

Necessary assumptions (17) and (18) on the additional unobserved state variables are not satisfied for already existing latent variables h_t and m_t introducing second type of unobserved variables in the model. Indeed, conditional independence assumption is violated, first, by the fact that health process evaluates conditional on the previous own value which are not recovered from the observations, besides, evaluation of the observable state variables (for example, labour market position) does in fact depend on the realization of both health and matching processes.

For careful implanting additional unobserved variables into the general model setup, state space must be expanded by Cartesian multiplication with $R^{|D|}$, choice sets redefined so that $D_t(ps_t, h_t, m_t, e_t, \varepsilon_t) = D_t(ps_t, h_t, m_t, e_t)$, probability distribution function $q(\varepsilon_t \mid s_t)$ defined so that $\Pr\{q(\varepsilon_t[d_t] \mid s_t) = 0\} = 1$ for all $d_t \notin D_t(s_t)$, instantaneous utility redefined according to (17) and transition probabilities redefined according to (18). Once this is done, theorems 3.1, 3.2 and 3.3 from (Rust, 1994) can be applied to state under minor additional technical requirements on the density $q(\varepsilon_t \mid s_t)$, utility function $u(d_t, s_t)$ and probabilities $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ the following result.

There exists a "social surplus" function

$$G_t\left(\{u(d_t,s_t),d_t\in D_t(s_t)\}\,|\,s_t\right)=\int\limits_{R^{|D|}}\max_{d_t\in D(s_t)}\left[u(d_t,s_t)+\varepsilon_t[d_t]\right]q(\mathrm{d}\varepsilon_t\,|\,s_t)\,, \tag{19}$$

which is concave in $\{u(d_t,s_t),d_t\in D_t(s_t)\}$ and has useful additivity property

$$G_t\left(\{u(d_t,s_t)+\alpha,d_t\in D_t(s_t)\}\,|\,s_t\right)=\alpha+G_t\left(\{u(d_t,s_t),d_t\in D_t(s_t)\}\,|\,s_t\right). \tag{20}$$

In the expanded sequential decision problem the optimal decision rule $\delta^*=\left(\delta^*_{T_0}(s_{T_0},\varepsilon_{T_0}),..,\delta^*_T(s_T,\varepsilon_T)\right)$ is given by

$$\delta^*_t(s_t,\varepsilon_t)=\underset{d_t\in D_t(s_t)}{\arg\max}\{v_t(d_t,s_t)+\varepsilon_t[d_t]\},\ t\in\{T_0,..,T\}, \tag{21}$$

where $v_t(s_t,d_t)$ is defined as (recall that S denotes the state space)

$$v_t(s_t,d_t)=\\ =\begin{cases}u(d_T,s_T)+\Lambda\left(s_T\right),\ t=T,\\ u(d_t,s_t)+\rho_t\beta\sum\limits_{s_{t+1}\in S}G_t\left(\{v_{t+1}(d_{t+1},s_{t+1}),d_{t+1}\in D_{t+1}(s_{t+1})\}\,|\,s_{t+1}\right)p(s_{t+1}\,|\,s_t,d_t),\ t<T.\end{cases} \tag{22}$$

Observed partial stochastic process $\left\{\tilde{d}_t,\tilde{s}_t\right\}_{\delta^*}$ induced by the optimal decision rule δ^* within the expanded sequential decision problem is Markovian with non-stationary transition probabilities

$$p_t(s_t,d_t\,|\,s_{t-1},d_{t-1})=P_t(d_t\,|\,s_t)\cdot p(s_t\,|\,s_{t-1},d_{t-1})\,, \tag{23}$$

where $P_t(d_t\,|\,s_t)$ is a well-defined (due to (20)) probability distribution

$$P_t(d_t\,|\,s_t)=\frac{\partial G_t\left(\{v_t(d_t,s_t),d_t\in D_t(s_t)\}\,|\,s_t\right)}{\partial v(d_t,s_t)}. \tag{24}$$

Combining (23) with (24) gives probabilities that can be used for likelihood function construction. Namely, under assumption that $q(\varepsilon_t\,|\,s_t)$ takes a particular form of multivariate extreme value distribution, when the components of the vector ε_t are independent and distributed identically with extreme value (Gumbel Type I) distribution,

$$q(\varepsilon_t\,|\,s_t)=\prod_{d\in D(s_t)}\exp\{-\varepsilon_t[d]+\gamma\}\cdot\exp\{-\exp\left(-\varepsilon_t[d]+\gamma\right)\},\ \gamma=0.577\,, \tag{25}$$

(24) has an analytical solution and $P_t(d_t\,|\,s_t)$ is very conveniently given by

$$P_t(d_t\,|\,s_t)=\frac{\exp\{v_t(d_t,s_t)\}}{\sum\limits_{d'\in D(s_t)}\exp\{v_t(d',s_t)\}}\,, \tag{26}$$

where $v_t(s_t,d_t)$ in this case is defined by

$$v_t(s_t,d_t)=$$
$$=\begin{cases} u(d_T,s_T)+\Lambda\left(s_T\right), t=T, \\ u(d_t,s_t)+\rho_t\beta\sum_{s_{t+1}\in S}\log\left(\sum_{d_{t+1}\in D(s_{t+1})}\exp\left\{v_{t+1}(d_{t+1},s_{t+1})\right\}\right)p(s_{t+1}\mid s_t,d_t), t<T. \end{cases} \tag{27}$$

Applying backward induction mechanism to formulas (27) and (26) allows successive calculation of $v_t(s_t,d_t)$ and probabilities $P_t(d_t\mid s_t)$, which can be plugged into (23) to give transition probabilities for the observed realizations of stochastic process $\left\{\tilde{d}_t,\tilde{s}_t\right\}_{\delta^*}$. Given the panel of observations $\left\{d_t^a,s_t^a\right\}_{t\in\{T_0-1,..,T\},a\in\{1,..,A\}}$ where A agents are indexed with index a (Rust, 1994) constructs the likelihood function

$$L'(\theta)=\prod_{a=1}^{A}\prod_{t=T_0}^{T}P_t(d_t^a\mid s_t^a,\theta)\cdot p(s_t^a\mid s_{t-1}^a,d_{t-1}^a,\theta). \tag{28}$$

Thus, the overall logic of the presented approach is to introduce additional random state variables which enter the optimal decision rule δ^* in (21) shifting the maximum point around the choice set so that the observed variables can not perfectly predict it, and then to integrate them out with the means of (19) deriving expressions (22-24) that characterize the observed partial state-decision process $\left\{\tilde{d}_t,\tilde{s}_t\right\}_{\delta^*}$. Assuming a particular form of multivariate extreme value distribution (25) for the introduced noise gives simple analytic expressions (26) for probabilities $P_t(d_t\mid s_t)$ allowing to avoid numerical integration in the computational procedures when calculating the likelihood function with the means of backward induction based on the formulas (26) and (27).

In the current model, however, there is a second group of unobservable state variables, namely h_t and m_t, which makes it impossible to implement the described procedure in a straightforward matter because of the lack of the data for calculation of (28). To deal with this I use the described approach once again and integrate these unobservables out of the likelihood function as well. This appears to be feasible because variables h_t and m_t constitute stochastic Markov process $\{\tilde{h}_t,\tilde{m}_t\}$ that evaluates independently of all the rest of the state and decision variables and has simply parameterized transition probability matrix composed of (15) and (16).

Denote $HM_t(ps_t,e_t,d_t)\subset\{0,1,2\}\otimes\{0,1,2\}$ a set of pairs (h_t,m_t) consistent with the other state variables and the decision in a given period. Consistency simply implies

$$\forall(h,m)\in HM_t\;\; p\left\{ps_t,h,m,e_t,sp_t,nw_t,aw_t,d_t\mid s_{t-1}\right\}>0. \tag{29}$$

In other words, set $HM_t(ps_t, e_t, d_t)$ contains all possible values of h_t and m_t variables that could be seen on the realizations of $\{\tilde{d}_t, \tilde{s}_t\}_\delta$ process. By the model setup health and match variables are not completely recoverable from the observations, but may be to some extent revealed by the labour market state as follows from Table 1. To see this, the table should be followed from left to right. For example, if the current labour market state is partial disability with employment in the non-AFP company ($ps_{t+1} = 5$), it can be seen that the only consistent health is $h_t = 1$. Then likelihood function (28) should be modified so that the probability mass from all realizations $\{(h_t, m_t) \in HM_t(ps_t, e_t, d_t)\}_{t\in\{T_0,..,T\}}$ of the $\{\tilde{h}_t, \tilde{m}_t\}$ process consistent with the observations of the rest of the state variables should be included.

Another adjustment of the formula (28) is needed due to the fact that, as it will be discussed in section 3, the panel used for estimating the model is unbalanced, so observations of different individuals cover different time periods. This small adjustment is achieved by introducing agent specific time indexes $T_0^a \in \{T_0,..,T\}$ and $T^a \in \{T_0,..,T\}$. When $T_0^a = T^a$, only single observation besides the initial one is available for a given household. Since the latent variables h_t and m_t are not observed in the initial observation, I make the following unified assumption.

First, $h_{49} = 0$, thus forcing all the agents be healthy one period prior to the $T_0 - 1$ when initial conditions begin to be recorded. This assumption is driven by two considerations. Initial condition may be somewhat related to the birth of an agent when it is reasonable to assume good health. Besides, as it will be discussed in section 3, I concentrate on the sample of agents who are active on the labour market at age 50, thus reducing their chances to be unhealthy at this age. However, there are some unhealthy people in the sample combining job and disability already at age 50 as their exclusion would bias the estimates of the transition probability matrix for h_t. Therefore it is important to allow for some health deterioration between the healthy age and 50. Thus, the initial condition for health at the initial period $T_0 - 1$ is random with a given probability distribution. When the agent specific initial period is different from 50, $T_0^a > T_0$, the corresponding distribution of $h_{T_0^a - 1}$ can be calculated in a standard procedure for Markov chains. Denote $\bullet[i, j]$ an element of a matrix in the row *i* and column *j*. Then since the health process is completely independent, the probability distribution for $h_{T_0^a}$ is given by

$$p_0(h_{T_0^a}) = \left(\pi^{(h)}\right)^{T_0^a - T_0 + 1}[0, h_{T_0^a}]. \tag{30}$$

Formula (30) also includes the case when $T_0^a = T_0 = 50$.

Similarly, the initial condition for matching process $\{\tilde{m}_t\}$ is also random. Calculation of the conditional probability distribution $p_0(m_{T_0^a-1} \mid h_{T_0^a-1})$ for given $h_{T_0^a-1}$ and the joint distribution of $(h_{T_0^a-1}, m_{T_0^a-1})$ is a little more complicated due to the fact that matching depends on health as it follows from (16). Three cases must be considered: for $h_{T_0^a-1}=0$ the distribution for $m_{T_0^a}$ is defined similarly to (30), for $h_{T_0^a-1}=1$ all possible times when health could have shifted from 0 to 1 have to be considered and the full probability calculated, and for $h_{T_0^a-1}=2$ $m_{T_0^a}=0$ by the assumption made previously. Similarly to health some value of match variable has to be fixed at early age (maybe $m_t=0$ just before the start of career).

However, the following considerations allow me to avoid these complicated calculations. As mentioned before, the sample of agents to be defined in section 3 includes individuals who are active on the labour market at the initial age (at T_0^a). From Table 1 it can be seen that m_t is recoverable for all these states implying that corresponding $HM_{T_0^a}$ set will only allow for one value of $m_{T_0^a}$ making the calculation of the corresponding probability distribution unnecessary. Therefore the initial probability distribution for matching is degenerate and

$$p_0(m_{T_0^a-1} \mid h_{T_0^a-1}) = 1. \tag{31}$$

Finally, if $\left\{d_t^a, ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\right\}_{t\in\{T_0^a-1,..,T^a\}, a\in\{1,..,A\}}$ is available data, denote HM^a the induced set of consistent trajectories $(h,m)^a = \left\{(h_t, m_t)\right\}_{t\in\{T_0^a,..,T^a\}}$ of the health-match process $\{\tilde{h}_t, \tilde{m}_t\}$ so that

$$(h,m)^a \in HM^a = \otimes_{t=T_0^a}^{T^a} HM_t(ps_t^a, e_t^a, d_t^a)\,. \tag{32}$$

In other words HM^a is a set that contains all trajectories of the health-matching process denoted $(h,m)^a$ consistent with the observed evaluation of the state variables for a given agent. Note that the notion of the consistent set $HM_t(ps_t^a, e_t^a, d_t^a)$ defined in (29) is used here $T^a - T_0^a + 1$ times to create a "corridor" which contains all possible values of health and match that could be met on the realizations of the health-match process consistent with the observed data.

The likelihood function for the model then takes the form

$$L(\theta)=\prod_{a=1}^{A}\left[\sum_{(h,m)\in HM^a} p_0(h_{T_0^a-1},m_{T_0^a-1},\theta)\prod_{t=T_0^a}^{T^a} P_t(d_t^a \mid s_t^a,\theta)\cdot p(s_t^a \mid s_{t-1}^a,d_{t-1}^a,\theta)\right]$$

$$p_0(m_{T_0^a-1},h_{T_0^a-1},\theta)=p_0(h_{T_0^a-1})\cdot p_0(m_{T_0^a-1} \mid h_{T_0^a-1}), \tag{33}$$

$$s_t^a=\left(ps_t^a,h_t,m_t,e_t^a,sp_t^a,nw_t^a,aw_t^a\right),$$

where the parameter vector θ includes already mentioned parameters $(\beta,\pi^{(h)},\pi^{(m)},\theta^{bhm})$ and some other defined later when empirical specification of the model is complete.

The constructed likelihood function (33) differs from (28) in additional summation over all trajectories of the latent health-match process $(h,m)^a$ from the set HM^a of all such trajectories consistent with the observed data. Summation is weighted with the initial condition probabilities $p_0(m_{T_0^a-1},h_{T_0^a-1},\theta)$ defined in (30-31). Note that the values of h_t and m_t in the observed state vector s_t^a come from these consistent trajectories and not from the observations as in (28). In other words the likelihood function is accounting for the probability mass from all possible realizations of the latent process, and thus integrating over the unobservables.

This approach could be regarded as incomplete information likelihood and corresponds well to the simulated likelihood which uses simulated sequences of unobservables to establish the likelihood calculation. Simple structure of the unobserved process allowed me to take into account all its possible realizations instead of limited number of simulated ones. (Rhenius, 1974) provides theoretical foundation for this approach showing the equivalence of incomplete information Markov decision models to their counterparts reformulated in terms of probabilistic distributions of unobservables. At the same time, this method is related to the well established EM algorithm (Dempster et al., 1977), which suggests iterating the expectation step when the latent variables are integrated out of the likelihood function conditional on the parameters of their distribution, and the maximization step when the optimal parameter values are found. In the current model, the algorithm collapses to just one joint step because of the simple distributional assumptions of the latent variables which appear to be dependent on the separable parameters.

Expression (33) concludes the description of the model and I move on to discussing the issues related to the data collection.

3. *Analysis of the data*

This section focuses on construction of the panel dataset $\left\{d_t^a, ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\right\}_{t\in\{T_0^a-1,..,T^a\},a\in\{1,..,A\}}$ used in calibration and estimation of the model. It summarizes variable collection techniques and presents adopted mechanisms for generating aggregated and stratified values. Most important of these are the definitions of labour market states which will be in focus throughout the paper. This section is started with strict sample description and goes through all observed variables giving references to some descriptive statistics in the Appendix.

3.1. Sample definition and data sources

The primary data source in the paper is the data collection at the Ragnar Frisch Center for Economic Research at the University of Oslo. This data sets are based on the administrative register files processed by Statistics Norway. The data sets give socio-economic information on the full Norwegian adult population mainly between 1992 and 2003 with wage histories traced back to 1967. Variables from different files and time periods may be directly linked on the individual level via an encrypted identification number. The whole panel thus covers demographic and family characteristics, annual employment data with recorded salaries and unified employer identification, data on social benefits including disability and old age pensions, unemployment registers constructed on monthly bases, etc. A broad description of the Frisch Center data collection can be found in (Hernæs et al., 2000).

With this rich data available I have to focus on a certain sample of households to be used. I adopt the following sample definition.

Given the observation window of 1992 to 2003 and the modeling period of ages 50 to 70, and trying to include in the sample longest possible observation sequences, I construct the sample from individuals born between years 1933 and 1942. Thus, the shorter observation windows is ensured to be contained in the longer modeling period leading to the unbalanced panel where the individuals are observed in 12 consecutive years and the middle part of the modeling period has highest numbers of observations (See Figure 12 in the Appendix, p. 102).

Although some data in the registers are recorded and updated monthly, the essential employment affiliation is only available on an annual basis, namely for calendar years. The absence of more precise records of timing drives the rather imprecise but inevitable authentication of calendar years with the years of particular age. Given this assumption, I disregard events within a year and "mark" time periods (calendar years) with the age a particular individual turns in a given calendar year regardless on the exact birthday date. Herefrom the age is the only time index.

To construct households the defined individuals are accompanied with the spouses who unlike the initial sample need not be born in the fixed time interval and therefore may not have as rich information from the available data. The families are constructed on the bases of the family register which contains both registered marriages and unregistered cohabitations when the couple has at least one common child. Some individuals turn out to be single, thus giving rise to two types of households in the model. In the first type, single households, the only person is a decision maker whose labour market behavior is modeled. In the second type, full households, the decision maker is assumed to be the initially selected spouse (born within given time limits) or if both spouses qualify, the main earner in the family (in the improbable case of exactly equal earnings, the male). For the rest of the paper the assumed decision maker will be referred to as the individual, the agent or the primary spouse and the second member of the family in case he or she exists will be referred to as the spouse or the secondary spouse. Existence of the spouse is marked with sp_t^a variable and it's motion will be discussed shortly. Gender of the primary spouse (also in single households) is one of the constants and will be referred to as gender of the household.

On the bases of demographic and family registers from 1993 the initial sample conditions define 106 452 single households and 200 162 full households with 71 327 primary wives and 128 835 primary husbands, altogether 306 614 households (see Table 18 in the Appendix, p. 94).

Further definition of the sample is due to initial conditions. To concentrate solely on transitions out of the labour market I require an active labour market position in the first year of observation for every individual. The notion of active labour market state will be described thoroughly in the next section, but in general it implies either working at least part time or looking for job as a registered unemployed.[r]

Finally, to take care of the outliers (whose labour market behavior is not likely to be governed by the economic model developed in the paper) I also introduce a simple earnings test for the before tax household income which should lie within the limits of 40 thousand and 1 million Norwegian crones (NOK) in 1992 prices.

The latter requirements reduce the number of households in the sample by 34.47% to 200 921 households mainly due to initial state requirement (69.98% of the reduction). The full report of the reasons for sample reduction is given in Table 19 in the Appendix (p. 94).

---

[r] For precision it should be noted that the initial condition is constructed from not one, but two first observations, because the previous labour market state is recordered with lag 1. Thus, only the sequences of three consecutive annual observations and more are taken into the sample (the former providing just one data point to be used in the estimation).

With the given definition the sample naturally combines several major household types which will have slightly different intentions for the labour market transitions. For the full households these are families with older primary husband who will be approaching retirement earlier than the spouse and whose decision to exit labour force will be captured by the model, and families with the younger primary wife whose older husband is simply unobserved within the available data. In addition these two groups each contain families of same age spouses among which the focus of the model is on the career oriented one (family earner). Single households contain both never married men and women and widows and widowers. The sample is thus characterized by considerable degree of heterogeneity which is hard to eliminate without considerable reduction of the general scope of the model or complicating the computational part by introducing additional control variables.

3.2. Labour market states

This section deals with labour market states construction which is the central challenge in the data analysis section. The general theoretical requirement is for the set of labour market states to be an exhaustive set of mutually exclusive alternatives, so that at all points in time all the agents have to be distributed across labour market states and no agent is allowed to occupy more than one state at a time. Even this simple requirement may be hard to satisfy with the real data.

Labour market states are constructed on the individual level by looking for the relevant records in a number of registers containing information of social security payments such as early and normal retirement pension, disability and unemployment benefits, etc. Such procedure is justified by the strong logical tie between a given labour market state and the main corresponding source of income. A slightly different definition is adopted for employment states because the mentioned link to wage income is much weaker as a consequence of greater heterogeneity among the employed. The link to time allocation and the trade-off between work and leisure seems to be another important identification source for employment and therefore should be utilized.

The main source of employment data is the employers register (AT) which holds annual records of worked hours and earned wages on the individual level. Statistics Norway links this register with the tax office database (LTO) to create a correspondence between the annual wage reported by the employer and those reported to the tax authorities by the individual. This joint file (ATmLTO) presents much cleaner data and is therefore used instead of AT register. Due to the way it is constructed, the joint file contains employment affiliation variables both as recorded by the employer and as recorded by the tax authority. Besides the wage data both of the agencies construct several time variables: first and last dates of the employment spells within a year, total number of days worked, an aggregated indicator for full time and part time employment[s]. Such large number of variables reflecting the same practical issue requires careful interpretation for each of them and gives rise to certain concerns whether the relevant interpretations were kept in mind by all the people maintaining the register. In particular, the great number of workers only registered with a particular job for one day in a year (mostly 1st of January or the 31st of December) suggests some "book keeping" when filling out these data base fields.

Simple analysis indeed shows that different variables on time of employment are poorly coordinated. For example, while the employment duration calculated from the start and finish dates from the AT register slightly overestimates the corresponding total days worked variable from the same register (with the R^2 in the corresponding regression at 96.22%), just the opposite is true for the LTO register (also with lower R^2 – 91.33%). The fraction of full time workers in the AT register (64.29%) is way lower than that of LTO register (76.40%). Finally, the correlation coefficient between the time of employment calculated from the two registers is only 0.3642. The overall impression is that AT data is more organized and seems to have less artifacts from the dates being filled out normatively.

Given unreliable data on the employment durations I decided to use multiple criteria to identify employment. One criterion is set to be a sufficient wage (over the basic pension G) in a given year, the other criterion is a sufficient (more than a month) duration of employment as measured by the AT register, and there is the third criterion that comes from a different register dealing with pension point accumulation. This last register recalculates annual wages in terms of pension points with some truncation and thus presents enough information to recheck the first criterion from the independent information source. This complex procedure allows construction of employment labour market state with more precision whereas low thresholds allow capturing partial employment as well as full time employment as intended by the model.

Technically, the labour market states are constructed from the results of the following four tests applied to each individual in every year when observation exists.

1. Test for employment: being listed on the employment register and either having more than 30 registered working days, having registered pension point greater than one (e.g. having annual earnings greater than one minimal pension in the corresponding year) or having annual wage greater than 36 000 NOK (in 1992 prices).

[s] Correspondingly ST_REG, SP_REG, AT_TOT, FORV_ARB and FRAMMDD, TILMMDD, ANT_DAG, ANS_KODE for employer and tax authority registers.

2. Test for unemployment: being listed on the unemployment register for more than 6 months within a particular year.
3. Test for disability: being listed among the receivers of the disability pension.
4. Test for pension: being listed among the pensioners on either old age or early retirement pension registries.

The fractions of all observed combinations of the individual test scores (either 1 or 0 for passing or not passing) represented by the four digits binary numbers are plotted against time on Figure 13 in the Appendix (p. 102). Clearly, the data displays much richer collection of combinations of the test scores than is desirable for clean definition of the 8 values of the state variable ps_t^a as introduced in section 2.3. In other words, it is clear that the tests are many times passed together indicating that the corresponding labour market states do not express the mutual exclusion property. Moreover, some of the observations have no scores which violates also the exhaustive property.

The latter problem is due to the absence of some individuals in the considered registers. These may be housewives or self-employed or other individuals not active on the labour market. To deal with the problem I introduced a residual out of labour market state (OLM) which consolidates the individuals with no scores on the given four tests and corresponds to $ps_t^a = 0$. Introduction of such a labour market state may be quite troublesome because the heterogeneity is high among the agents in this state, but as it will be seen below this can be dealt with.

The former problem is mostly due to the assumed annual timing structure of the model while in reality the labour market state may change within a year. There may also be errors and misinterpretations in the registers. It can be noted from Figure 13 however that the unreasonable combinations of the test scores are rather uncommon. In these circumstances I have decided to pursue the labour market definitions suggested by the data without trying to clean things up with a prioritizing rules that would allow identifying the "main" state on the labour market for each individual[t]. This approach identifies the most common labour market states and merges smaller groups into the main ones. The number of the identified states is determined by the degree of computational complexity the model can bear.

The main identified labour market states were already presented in the description of the state variables in the section 2.3, while the aggregation principles used in the definitions are spelled out below.

---

[t] One example of such rule would be the source of maximum income identifying a labour market state.

- All individuals are forced to pension at age 70. This is justified by the institutional rules and deviations may be regarded as errors in the registers.
- Test 4 is disregarded before the earliest retirement age of 62[u]. However, the pension record is given strict priority over the rest of the records after age 62.
- Unemployment may include partial employment lasting for less than a half of a year (as follows from test 2).
- Probable spells of unemployment for those partially disabled and working are disregarded.
- Disabled may look for a job and when finding one are classified into the corresponding labour market state. However, when not successful they are occupying disability labour market state instead of unemployment.

Table 2 reflects enumerated principles and summarizes the definitions of the labour market states by presenting correspondence of 4 digit binary codes for test scores to the values of the ps_t variable.

Table 2. Definitions for the labour market states ("•" denotes either 0 or 1).

Labour market state		Test scores				Additional conditions
ps_t^a	Name	1	2	3	4	
0	OLM	0	0	0	0	age < 70
1	Pension	0	0	0	1	-
		•	•	•	1	age > 61
		0	0	1	0	age > 66
		1	0	1	0	age > 66
		0	1	0	0	age > 66
		0	1	1	0	age > 66
		•	•	•	•	age = 70
2	Disability	0	0	1	0	age < 67
		0	•	1	1	age < 62
		0	1	1	0	50 < age < 67
3	Unemployment	•	1	0	0	age < 67
		•	1	0	1	age < 62
		0	1	1	0	age = 50
4, 6	Employment	1	0	0	0	age < 70
		1	0	0	1	age < 62
5, 7	Employment with partial disability	1	•	1	0	age < 67
		1	•	1	1	age < 62

[u] Pension register however does list a small number of people retired earlier than that on some special conditions (firefighters, dangerous working conditions, etc.)

The additional distinction of AFP employment is performed on the bases of the dynamic list of AFP companies which construction is based on tracking the last place of employment for the actual early retirees listed in the corresponding AFP register. Tracked company is assume to have joined the AFP program one year prior to the first observed early retirement of it's staff (details can be found in (Iskhakov and Kalvarskaia, 2003)).

The described procedure ensures that the resulting set of labour market states satisfies requirements of the well defined set of alternatives in every time period. Observed fractions of individuals in each of the labour market states are presented by age in Figure 14 in the Appendix (p. 103). The plot illustrates the processes taking place among the elderly workers on the labour market. Employment is the most occupied state up to early 60s ages, and is then overtaken by the two major causes of leaving the labour force – disability and early retirement (which increase from 62 is accompanied with the sharp decrease in AFP employment). Unemployment which is allowed by the initial conditions and has second biggest portion of the sample at age 50 also gradually decreases up until 67 when the unemployed are forced into old age pension. Full time disability increases in a concave formation which is also observed in the partial disability-employment states. The OLM state has increasing share of the sample up to age 67 and then sharply decreases. Overall the behavior of the elderly workers matches the expectations and the logic that drove the development of the structural dynamic model.

Contrary to the aggregated view of fractions of the sample in particular labour market states and their dynamics Figure 15 to Figure 19 (pp. 103-105) present the revealed patterns of transitions among the labour market states on the individual level. The figures present 20 transition matrices between every pair of ages from 50 to 70. Labour market states are coded with the values of ps_t variable introduced in the previous section with additional value 8 indicating death. The dots set on the crossings of the gridlines indicate transitions between the corresponding states at given ages, sizes of the dots are proportional to the number of agents making corresponding transition. The graphs visualize the process of transition of the elderly workers out of the labour market which can be presented as the box combining labour market states 3 to 7 on every axes. By age 67 most of the individuals are retired whereas by age 70 the only occupied columns are pension and death. Absence of transfers in the upper-right region of the matrices justifies both the absorption assumptions made earlier for the passive labour market states (0 to 2) as a group and strict absorption of pension state.

Thus, the labour market states constructed from the register data satisfy all the requirements and provide the necessary observations of ps_t variable.

3.3. Demographic dynamics

There are three processes of demographic dynamics embedded into the model: death of the primary spouse (in both single and full households), death of secondary spouse (in full households only) and the break up of the full household not due to the death of a spouse, apparently divorce. The first process is easily modeled with standard survival probabilities. The second process is much harder to model within the current set up because the only time index used in the model for practical purposes is the age of the primary spouse and therefore standard survival probabilities can not be used for the spouse. I go around this by constructing spouse survival probabilities on the available sample using the only available time index. The third process is the most difficult to embed into the model and requires special investigation.

Table 3 Family dynamics: starting and finishing years of the observation spells and corresponding numbers of families observed.

Frequency Percent	1993	1995	1996	1998	1999	2000	2001	2002	2003	Total
1993	7282 3.42	7271 3.42	3585 1.69	3575 1.68	3739 1.76	3984 1.87	3957 1.86	4233 1.99	163206 76.72	200832 94.41
1994	0 0.00	550 0.26	243 0.11	212 0.10	237 0.11	217 0.10	196 0.09	198 0.09	3046 1.43	4899 2.30
1996	0 0.00	0 0.00	97 0.05	125 0.06	116 0.05	97 0.05	88 0.04	92 0.04	1878 0.88	2493 1.17
1997	0 0.00	0 0.00	0 0.00	58 0.03	48 0.02	49 0.02	31 0.01	37 0.02	683 0.32	906 0.43
1999	0 0.00	0 0.00	0 0.00	0 0.00	58 0.03	53 0.02	43 0.02	38 0.02	638 0.30	830 0.39
2000	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	54 0.03	53 0.02	49 0.02	717 0.34	873 0.41
2001	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	45 0.02	39 0.02	637 0.30	721 0.34
2002	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	48 0.02	578 0.27	626 0.29
2003	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	542 0.25	542 0.25
Total	7282 3.42	7821 3.68	3925 1.85	3970 1.87	4198 1.97	4454 2.09	4413 2.07	4734 2.23	171925 80.82	212722 100.00

Table 3 presents a short numerical analysis of family fluctuations based on the family registers from 1993-2003. A value in a row X and column Y shows the number of families observed continuously from year X to year Y (inclusive). There are several reasons for different cells in a table to be non-zeros, namely, two reasons for a family to appear – either after a marriage (samboer) registration or first appearance in the observation period of 1992-2003 – and three reasons for a family to disappear – divorce, death of a spouse or moving out of observation period. By checking whether the spouse actually died in a given year (this information is available from the demographic files) and disregarding the observations slipping out of the data window it is possible to calculate sample specific divorce rates in every time period.

The described processes of demographic dynamics are presented in Figure 20 in the Appendix (p. 106). As expected, the sample specific death rates are greater for males that for females, but differences along primary–secondary dimension are quite interesting. While for females the difference between primary and secondary spouse are neglectable, death rates for secondary males are higher than those for secondary females. This straightforwardly corresponds to the typical households mentioned earlier: when males are secondary in the households, this is very likely because they are older and slip out of the data window. But the death rates are calculated with respect to primary spouse age, therefore they appear much higher than for the primary males. The fact that this effect is not visible for females simply reflect the tendency of age equality in the full households with primary husband.

Since the model is built around only the primary spouse, the events of death of the secondary spouse and divorce are undistinguishable. The only traceable event with this respect is shifting of the sp_t variable from 1 to 0. Therefore it becomes possible to combine these two probabilities into one for the use in sp_t evaluation. Combined death of spouse and divorce rates are also plotted in Figure 20. It is clear that the drastic difference in the combined probabilities is mainly due to the differences in the spouse survival discussed above. The divorce rate is rather independent of the gender of the household and although quite high compared to the death rates, much lower than the all ages average which is expected.

Under earlier assumption of no new marriages discussed probabilities are enough to characterize the processes of deaths and divorces incorporated into the model and represented by the variable sp_t.

3.4. AFP eligibility

According to its definition the state variable e_t keeps track of the otherwise non verifiable individual criteria for early retirement pension. Available data allows for reconstructing AFP eligibility thoroughly according to the institutional settings described in section A.1 in the Appendix[v]. Namely, the following checks are made:

- Last two annual wages are over the basic pension amount G.
- At least 10 years after age of 50 have annual wage over the basic pension amount G.
- The average of the best 10 annual wages throughout working life is over 2G.
- Individual is employed in the AFP companies for at least last 3 years.

---

[v] To be precise there is one exception, namely the rule that an individual is not receiving any payments from the employer other than labour related.

The last condition is used instead of more complicated one in the actual AFP rules that requires employment in the same AFP company for the last 3 years or employment in different AFP companies for the last 5 years. Tracking the change of employers is somewhat tricky with the available data because it appears very hard to reliably identify same employer in different years, therefore the actual rule was to some extent simplified.

Altering the AFP rules for eligibility reconstruction did not skew the aggregated image of the AFP program coverage. Table 20 in the Appendix (p. 36) presents the fractions of the AFP eligible workers by age which correspond well to the overall figures. As it was set in the model description, e_t variable is only defined at the periods of possible early retirement, therefore only the middle section of the table (ages 62-66) is used in the model estimation.

3.5. Income

A long tradition in the labour economic literature appeals to the use of consumption and leisure as two major variables in the utility function, maximization of which under the budget constraint resolves the trade off between higher consumption accompanied by more working hours and higher leisure. Keeping this approach in the current model is highly desirable but rather hard. Previous section showed that the available data on worked hours is unreliable and the consumption information is not available from any register sources. In these circumstances several studies (Dagsvik and Strøm, 1992; Hernæs et al., 2000; Hernæs and Strom, 2001; Hernæs et al., 2006) have used discrete leisure variable defined as a deterministic function of labour market state and income to approximate consumption. I apply the same principles below when defining deterministic part of the utility function (17) while this section is devoted to income analysis which takes on the main load of explaining the levels of utility that result from different decisions made by the agents.

First, I investigate how plausible it is to assume away savings. Life cycle theoretical framework suggests considerable movements in savings induced by the consumption smoothing effect at the end of working life when wage incomes are sharply decreasing. This effect may however be offset by the fact that large part of savings is likely to be locked in the illiquid durable goods, in particular housing. Figure 21 in the Appendix (p. 106) presents the dynamics of net household wealth averaged within 10 fixed groups representing deciles of lifetime averages. As it follows from the graph changes in wealth are very moderate and smooth in all parts of the whole wealth distribution except perhaps the tails. This is a clear indication of the fact that on average savings are quite constant in the considered sample and therefore may be assumed not to play important role in the explaining labour market transitions. Still, some savings related processes like for example bequest motives are captured by the model in the termination function $\Lambda(s_T)$ to be defined later.

Comparison of the bottom graph in Figure 21 with Figure 12 indicates that the available wealth data is limited and prevents constructing of consumption variable from subtraction.

One of the hardest challenges in developing models of discrete choice in the McFadden's conditional sense, that is when the decisions are to be affected by the characteristics of the alternatives (through utility function), is the need to construct these characteristics for the alternatives not chosen by the decision maker. Indeed, the data usually provides information only on those labour market states actually occupied by the agents and gives no information on, for example, income in other labour market states. There are two ways to deal with the problem. First, some simplifying assumptions may be made with respect to these unknown characteristics of the alternatives that allow their direct assessment. In the income example this approach would lead to careful thorough calculation of e.g. potential early retirement pension from the available wage histories and other data for the people who could have retired but chose not to. But as it is impossible to take into consideration and directly calculate other minor sources of additional benefits possibly existing in reality, they have to be assumed away. Moreover, the dynamic model requires the state variables to be self-sufficient in predicting the next period values. Together with the described technique this would imply that all variables such as income histories relevant for accurate calculations would have to be part of the state space in the model, which is clearly infeasible because of complexity limitation.

Therefore I adopt second approach which suggests forecasting characteristics of the potentially chosen alternatives using statistical models estimated on the existing observations of the alternative characteristics included into the state vector. In other words, the agents occupying a given labour market state provide information on income in this state, and it is used for constructing a probabilistic relation to state variables observed in whole sample, which in turn can be used to forecast potential income in this labour market state for all agents. The presented approach may suffer from small numbers of available in each labour market state observations, but given the large sample it is not a problem in this paper. Another open question is the accuracy of predictions of these statistical models. Indeed, with the great degree of heterogeneity also associated with the broad sample it will appear to be a considerable issue. Nevertheless, inevitable simplifications in representation of the modeled processes together with inevitable computational complications otherwise induced by the first approach and the aspiration to preserve accuracy of income data provide grounds for the use of statistical approach and the need to investigate available income data.

The data vector $\left\{d_t^a, ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\right\}_{t\in\{T_0^a-1,..,T^a\},a\in\{1,..,A\}}$ contains two indicators of income (aggregate wage aw_t and the number of consecutive years with high incomes nw_t) which relate to its long term and short term behavior as was mentioned in the model description. This particular setup is not crucial for the model itself and presents a compromise between the accuracy and complexity of the model. Since leisure will be a discrete deterministic function of labour market state so that income is the only available continuous explanatory variable, I decided to introduce additional discrete short term income indicator which is used for calculating benefits in the social security system and therefore should be useful in explaining utility levels in the passive labour market states.

Variable nw_t is almost directly observed in the registers. Following the rules of the social security system, I adopted the basic pension amount as threshold to define sufficient incomes. It became therefore possible to use earning histories collected for the purpose of pension calculations to simply count the number of years with pension points over one while truncating it at 10 and bringing down to 0 every time lower pension point was detected[w]. Resulting simple pattern of evolvement for variable nw_t (at each iteration it can either become 0 or increase by 1 unless already at the truncation level) is made use of when discovering its motion rule in the next section.

Aggregate wage aw_t was not given any specific interpretation at the modeling stage to allow for several observed variables to be tried out to play its role. The goal was to find a variable that on one hand could be easily forecasted one year ahead with the means of the rest of the state variables deployed in the model, and on the other hand was capable of explaining income levels in different labour market states. Three candidates were considered: simple average of the annual wages, average over annual wages in the best 10 years of working life and average of the best 20 years. All three of these variables were constructed from the earnings histories expressed in pension points. Relevant amounts for the basic pensions and CPI indexes were used to enforce common unit of measurement of 1 000 NOK in 1992 prices[x]. The reasons for considering these three variables as possible aggregate wage is the following. All three present long term trends in individual wage earnings and therefore contain information on expected incomes in all labour market states. Average wage is probably least powerful in this respect since pension and disability benefits are calculated from the best averages. On the other hand, simple average evaluates according to deterministic rule while the other two variables need statistical models for its evaluation and the very desirable property of deterministic evaluation of the continuous state variable may not be obtained. Nevertheless, all three variables were calculated and their dynamics are presented in the

---

[w] Such calculations were started at age 40 to allow for maximum value of 10 in the initial period.

[x] With inevitable loss of information on annual earnings below 1G and above 12G due to truncation in pension points calculation as it follows from the social security rules.

Figure 22-Figure 24 in the Appendix (pp. 107-107). Here again annual means within 10 fixed groups representing deciles of overall averages are plotted against time. As follows from the plots all three variables display very smooth motion in all parts of the distribution. The latter two also naturally display non-decreasing behavior.

The rest of this section presents the analysis of incomes in different labour market states that have to be explained by the aggregate wage together with other state variables. Discussion of the statistical models developed in this respect and the presentation of the chosen aggregate wage variable is left for the next sections.

The study of the income sources is based on the data from several different registers. Major sources of income as wage, old age and AFP pensions and disability pension were investigated from the specific registers. In addition the database kept by the social security office containing information on all the rest of the benefits received by individuals was used in the investigation of secondary sources of income[y]. The latter database is organized so that paid amounts are accompanied by Norwegian tax codes (LTO) and therefore can be categorized into several groups that appear to be most important in the sample:

- Employment incomes that in addition to wages contain different occupational benefits provided by the employer (reimbursement for communication, commute expenses, etc.), holiday pay, travel allowance, etc.
- Pension incomes consist of different pension benefits: disability benefits up to age 67, possibly AFP benefits between 62 and 67 and regular old age pension benefits after the age of 67.
- Additional incomes combine the rest of the important sources of income for an individual such as occupational pension, survival benefit, unemployment benefit, child care, sickness and other forms of benefits resulting from governmental or private insurance schemes, additional personal old age annuities, golden shake premiums, capital income, etc.
- Spouse income comprises previous three groups calculated for the spouse.

Altogether these four sources of income allow for construction of total household income plotted in 1000 NOK in 1992 prices in Figure 25 in the Appendix (p. 108). The graph reflects the average decline of employment income with age which is accompanied with the gradual grow of pension income and additional income that include retirement related benefits. The spouse’s income is

[y] Correspondingly, ATmLTO, Afpp, Afpo, Aldp, Ufp and LTO Trygd.

slightly declining towards the retirement age and growing again with pension[z]. The very high replacement rate observed in household incomes should be taken cautiously because of the aggregation over all observations on one hand and the fact that the number of observations is falling towards both ends of the time interval on the other hand.

The four groups' composition of the income sources presents nice opportunity for assigning different types of income to different labour market states and in general to different combinations of the state variables. This way, spouse income should naturally be included into the total household income only when variable sp_t is 1 indicating a full household. The collection of figures in the Appendix (Figure 26 to Figure 33 on pp. 108-112) investigate the dependence of different income sources on the labour market states. Each figure contains two plots to display both the average level of the corresponding income group and the number of observations the average is taken over. The following revealed correspondence clearly justifies the definitions for the income groups and provides grounds for income forecasting in each pair of values (ps_{t+1}, sp_t) in the dynamic model:

- In the OLM labour market state only spouse's income and additional income constitute considerable income sources (See Figure 26).
- In retirement pension is the main income source while spouse's income and additional incomes play secondary roles. The sharply decreasing employment income is most likely to be an artifact of the assumed annual time periods used in the model and possible phased retirement disregarded here (See Figure 27).
- In the disability state as the number of observations grows, the stable image of the income structure is obtained. Most important is disability pension within the pension income group, also important are spouse's income and additional incomes (See Figure 28).
- In the unemployment labour market state the most important source is probably wage and employment income group which does not appear to be a controversy because unemployment is defined to include partial employment as well. Not least important although giving lower income is spouse incomes and additional incomes including unemployment benefit (See Figure 29).
- In the employment labour market states quite expectedly the employment income group is by far most important with somewhat higher wages available for employees of the AFP

[z] Spouse incomes appear relatively smaller that other sources of incomes since many observations represent single households and thus zeros appear in the calculation of the average.

companies. Spouse incomes also constitute considerable source while additional incomes are much more common for non-AFP employment (See Figure 30 and Figure 32).

- Combined employment and partial disability labour market states present the cases where all groups of income sources are of considerable importance although smaller in absolute value both compared (in the relevant groups) to full time employment and full time disability. The non-AFP case distributes the importance of wage, pension and spouse income quite uniformly while the AFP employment seems result in relatively higher wages which depart upwards from the rest of the income sources. In both cases additional incomes represent smaller but rather stable component (See Figure 31 and Figure 33).

This concludes the analysis of the income data. Short description of the decision variable in the next section precedes the further discussion on forecasting incomes, best aggregate wage and the specification of the utility and termination functions in the empirical specifications in the next section.

3.6. Decisions

By this time the data vector $\left\{d_t^a, ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\right\}_{t\in\{T_0^a-1,..,T^a\},a\in\{1,..,A\}}$ necessary for the model estimation is fully populated except the decision variable d_t^a which was not yet discussed. Recall from the model setup that the decision variable indicates the intentions of an agent to change his or her labour market state by providing answers to the questions "To remain on the labour market or not?" and "Apply for pension or not?". Since the only type of data in paper is register data, it does not provide any information on the intentions of the individuals.

The problem is solved by the fact that decision variable is completely recoverable from the observations of rest of the variables as it follows from Table 1. To see this, again the tables has to be followed from right to left: for each combination of state variables there is only one feasible value of the decision variable. This implies that as in the definition of the choice sets (9) the decision variable could be dropped from the likelihood function in (33). It would however be wrong to think of the model as too restrictive because of this property since it has nothing to do with the definitions of the choice sets (e.g. restricting the choices).

This concludes the section devoted to the data description and allows me to continue with definitions of the motion rules and other elements of the model in some way dependent on the data.

4. *Empirical specifications*

This section completes specification of the model giving functional forms for the utility and termination functions, completing definitions of the motion rules and describing additional elements of the utility. It starts with complete specification of the utility function and finishes with enumeration of the parameters of the likelihood function.

4.1. Preferences

This section specifies the functional form for the deterministic part $u(d_t, s_t)$ of the instantaneous utility function $U(d_t, s_t)$ (17) in the sequential decision problem of the agent (2).

As it was already announced in the previous sections, it is to be defined along the traditional lines as indirect utility dependent on income $I = I(t, d_t, s_t)$ and leisure $L = L(t, d_t, s_t)$. Following in addition a number of microeconometric studies on the labour market data I adopt the additive functional form with constant relative risk aversion (CRRA) with respect to household disposable income[aa] and linear leisure.

$$u(d_t, s_t) = u\left(I(t, d_t, s_t), L(t, d_t, s_t)\right) = a\frac{\left(Tx(I)\right)^{\lambda} - 1}{\lambda} + b(\overline{s}_t) \cdot L + \sum_{k=0}^{7} c_k \cdot \xi\left(ps_{t+1} = k\right), \quad (34)$$

where $Tx(\bullet)$ is a mapping from before tax household incomes to disposable household incomes, and $\overline{s}_t$ is extended state vector which contains all the state variables in s_t and in addition gender and individual early retirement age. A scalar coefficient a is used to scale the impact of the utility of income which is measured against the utility of leisure and additional non pecuniary utility generated by the different labour market states[bb] and weighted by the set of parameters c_k. A compound coefficient with leisure superposed as a function of the extended state vector captures the individual heterogeneity of preferences. Definition of the particular functional form for coefficient b is left for section 6 (where I discuss the estimation of the model) because the width of the reflected heterogeneity will apparently be defined by the computational tractability considerations.

Leisure is a deterministic function of the labour market state and is simply calculated as a fraction of time available for leisure after 8 hours of sleep per day and time spent at work. Active labour market states are assumed to occupy full 7.5 hours per working day (37.5 hours per week) except

[aa] CRRA as Box-Cox transformation also presents convenient flexible generalization that includes both linear ($\lambda = 1$) and logarithmic ($\lambda \to 0$) specifications.

[bb] Indicator function $\xi(\bullet)$ returns one if the condition is satisfied and zero otherwise.

the unemployment state which is assumed to occupy half of this time. OLM, pension and full disability labour market states have maximum leisure.

$$
\begin{aligned}
L_1 &= 1\text{-}\frac{37.5\cdot 52+8\cdot 365}{24\cdot 365}=0.444,\\
L_2 &= 1\text{-}\frac{\frac{1}{2}\cdot 37.5\cdot 52+8\cdot 365}{24\cdot 365}=0.555,\\
L_3 &= 1\text{-}\frac{8\cdot 365}{24\cdot 365}=0.667.
\end{aligned}
\tag{35}
$$

Thus, $L\in\{L_1,L_2,L_3\}$, and finally leisure function $L(t,d_t,s_t)$ can be precisely defined in terms of state variables of the model as

$$
L(t,d_t,s_t)=\begin{cases} L_1,\ ps_{t+1}\in\{4,5,6,7\},\\ L_2,\ ps_{t+1}\in\{3\},\\ L_3,\ ps_{t+1}\in\{0,1,2\}.\end{cases}
\tag{36}
$$

It should be noted that the given correspondence of leisure with the labour market states leads to the identification failure of the constant part of the leisure coefficient $b(\overline{s}_t)$ because it is not distinguishable from the labour market state specific coefficients in the third component of the utility function (34). Therefore, the constant will have to be dropped from $b(\overline{s}_t)$ in estimation.

Compared to leisure household disposable income needs a somewhat more complicated definition. Analysis of the income data in section 3 defines four major income sources for the household and reveals a certain correspondence of these source to the different labour market states ps_{t+1} as well as the spouse indicator sp_t. In this section I proceed with defining statistical relations of the different income sources to the model's state variables and present my choice of aggregate wage measure.

Several considerations were taken into account when estimating these quantitative relations. First, because of the limited number of explanatory variables in the state vector, the information had to be used to the full extent. The assumed timing convention allows using in the income equations both previous and current period labour market states because the utility calculation for the current period takes place after the state variables are known (previous labour market state among them) and current decision defining current period labour market state is taken. Thus, the model is capable of conditioning the income levels not only on the current labour market state, but on a corresponding transition on the labour market.

Second, even though the established correspondence reveals neglectable income sources in some of the labour market states, it does not work the other way around, and the problem of censoring at zero must be addressed. It turns out that wage, pension and spouse income are hardly affected by this: censored regressions estimates deviate very little from the ordinary regressions due to small number of censored observations. However, in the case of additional income this issue has an important influence. To preserve statistical properties of the residuals in the model for additional income (for the reasons that will be clear shortly) I decided not to use censored regression and instead estimated an auxiliary logit model that predicts positive outcome. The model estimated for positive additional income is then applied in forecasting for the individuals with favorable logit prediction.

Third, an appropriate approach for regression models selection had to be chosen. Goodness of fit measures and the estimated standard error of the residuals were the most important factors for forecasting reasonable income levels in different labour market states. These two criteria were best met with the models based on aggregate wage calculated as average of the 20 highest annual incomes up to a given age. Herefrom this definition of the aggregate wage is used throughout the estimation.

Table 3 presents the summary of the estimated income equations. Each of the equations was estimated on a specific dataset (defined by the "estimated on" filter) and intended to be used for specific individuals (defined by the "applied for" filter). These filters are set in accordance to the described correspondence of income sources and labour market states so that the income equations are estimated only on the data from those actually in particular states (working, pension, etc.) and forecasts are applied for people transferring to these states. This forecasting mechanism is not free from the possible selection problems, but bearing in mind the large number of observations and therefore implied aggregating over immense amount of data, I disregard the selection problems on the grounds of their comparatively smaller importance.

It should also be noted that the income sources are assumed to be independent and can therefore be forecasted separately under ordinary regression assumptions. This quite considerable simplification facilitates an innovative method for the calculation of intertemporal utility (explained below) and therefore allows for a considerable reduction of the computational load of the model and simultaneous maintaining a reasonable level of accuracy of the predictions.

All the coefficients in Table 3 are estimated with p-levels of the order 10^{-5} or less. Similar values obtained for F-tests are not reported. More detailed Table 21-Table 25 are available in the Appendix (pp. 95-98).

Table 4. Summary of models for different sources of income (ordinary regressions and one logit model marked with *).

	Dependent variable	wage income			pension				additional income		spouse income	
Estimated on (filter)	Transition (state codes)	(≥3)-(≥3)			1-1		2-2	•-(5,7)	any*		spouse exists	
	Age, other	<60	60-67	≥67	afp-67	≥67				inc>0	gender 0	gender 1
Applied to (filter)	Transition (state codes)	(≥3)-(≥3)			?-1		?-2	•-(5,7)	any		spouse exists	
	Age, other	<60	60-67	≥67	afp-67	≥67				logit>½	gender 0	gender 1
Variable		estimates										
Constant term		33.046	-28.438	-82.030	161.099	-1692.1	375.909	21.978	-0.747	25.993	-502.222	664.095
AFP age					-1.901	-7.880	-5.288				9.635	-8.726
Gender		16.474	13.339		-6.244	-15.981	-6.690	-8.595	0.325	12.928		
Time index (age-50)		-1.250				242.880		0.617	0.167	1.408	-8.608	-17.124
Time index squared			-0.166			-6.442				-0.102	0.069	1.080
Spouse indicator (*sp*)									-1.081	-5.121		
Number of high wage years (*nw*)		-2.080	-4.591			-4.964		-2.221	-0.093	-1.017		-0.601
Aggregate wage (*aw*)		0.229	1.580	2.240	0.532	0.534	0.403	0.243	0.0005	0.118	0.365	0.245
Aggregate wage squared (aw^2)		0.001	-0.007	-0.009	-0.001	-0.001	0.000			0.000	-0.001	
Aggregate wage to the third (aw^3)			9.9E-06	1.5E-05								
Cross effect of *aw*nw*		0.042	0.060	0.053				-0.010				
Prev. non-AFP employment								7.802				
Prev. partial DI + non-AFP employment								16.200				
Prev. AFP employment								20.509				
Prev. partial DI + AFP employment								16.365				
Current period OLM										63.230		
Current period pension retirement									0.346	-17.351	21.226	
Current period full disability									3.157	-21.772	10.907	
Current period unemployment									-1.042	9.563	9.828	
Current period non-AFP employment		48.070	46.185						-1.337		9.071	
Curr. partial DI + non-AFP employment		-18.853	-14.388						3.257	-29.762	11.589	
Current period AFP employment		51.332	56.137						-1.538		18.651	
Curr. partial DI + AFP employment		-8.728	-7.919					-7.979	3.728	-24.644	18.698	
Number of observations		914 839	703 107	11 296	53 212	108 727	216 503	151 419	2262566	886 269	1076041	433 340
R-square (pseudo for logit)		72.36 %	66.63 %	64.45 %	32.12 %	22.33 %	60.87 %	35.74 %	0.5088*	39.46 %	10.83 %	9.02 %
Estimated standard error of residuals		50.767	63.028	80.209	31.133	55.415	21.466	30.311		38.659	83.960	110.555
Number of coefficients		12	13	6	6	9	6	11	12	15	14	7

The large number of equations and coefficients presented in Table 3 (altogether 11 equations and 113 coefficients for 4 sources of income) reflects the complexity of the task of income prediction based on the limited information available through the state vector. Wage income is represented by three equations corresponding to three different age intervals: regular working careers, pre-retirement and exceptional employment after normal retirement age. These periods are chosen in accordance with the growing variance of the wage earnings revealed by a series of age-specific models (See Table 21 in the Appendix, p. 95). The first two equations are very similar and both assign the lowest average wage level to non-AFP employment with partial disability and the highest to the full time AFP employment whereas unemployment (including part-time employment) representing the reference group occupies the median level. Post retirement age employment is rather rare and is well explained by the lower levels of aggregate wage which leads to considerably smaller number of covariates for the model with comparable explanatory power.

Pension income falls into four distinct categories each of which is modeled with a separate equation (See Table 22 and Table 23 in the Appendix, p. 95). These categories (AFP pension, old age (NIS) pension, full and partial disability pension) can be clearly separated with the labour marker state and age variables facilitating the data fragmentation for the four regressions. The question however arises about whether the first retirement year is relevant and should be included into the dataset for pension equations estimation. Simple analysis shows that on average AFP and old age pensions are considerably lower in the first year of retirement than in the following years. This is probably due to discrete time nature of the model: I disregard possible within year fluctuations of income and therefore end up observing lower pensions for those retiring during the calendar year. I therefore drop the first retirement years from the datasets used in the estimation of the first three pension equations. The same is not possible for partial disability pension since the incomes in the first and following years do not differ systematically. Concentration on stable pension benefits in estimation of the pension equations is achieved by conditioning on the transfers from and to the corresponding labour market states. A consequence of this approach is the impossibility to use state dummies as explanatory variables and thus pension levels are mainly conditioned on the aggregate wage, time and fundamental household characteristics. In partial disability equation however, state dummies are used for both previous and current periods (their numbers are determined by considered transfers on the labour market) and display higher partial disability benefits to previously employed and currently employed in the non-AFP companies. The later indicates the fact that individuals with lower disability pensions are sorted into AFP employment.

Modeling of additional income is complicated by the fact that many households do not have any income from this source and the problem of censoring becomes quite considerable. I estimate a separate logit model for positive additional incomes which facilitates the use of the regression estimated only on the positive values of the dependent variable. Additional income is then calculated only for the households which produce a probability prediction over one half with the logit model. Both logit and ordinary regressions for additional income are displayed in Table 24 in the Appendix, p. 97. As it follows from the table female gender, absence of a spouse, bad employment situation in the recent years, disability, retirement and age positively affect the probability of additional income which is also higher for females and singles and those out of labour market.

Spouse income is modeled separately for two household genders[cc] reflecting the two major types of households (retiring older husbands with younger wives and retiring younger wives with husbands already on pension). In the former case, the current labour market situation of the leading spouse is used as an instrument for the spouse's situation whereas in the latter case this relationship disappears. The latter cases is the least accurate as the amount of information available allows for only taking simple averages over the groups of households with similar characteristics.

In general, analysis of goodness of fit in the models reveals the main complication of the chosen statistical approach to solve the problem of unobserved characteristics of alternatives. A large number of observations and considerable heterogeneity embedded into the data as well as the limited number of explanatory variables resulted in quite poor fit of the predicting models. As it follows from the last rows in Table 3 the best R-square of 72.36% is found in employment income equation while the most problematic ones are spouse's income, in particular in the households with primary wife (R-square is only 9.02%)[dd].

Poor fit of the regression models will cause inaccurate income predictions and thus can spoil the calculation of the utility in different labour market states which in the end will affect the calculation of the likelihood function and influence parameter estimates. However, this influence will not violate the asymptotic properties of the estimates because the errors are introduced only into the utility function and can be reformulated as its suboptimal specification. Therefore the only negative consequence of this inaccuracy in income prediction is reduced goodness of fit.

The problem can be dealt with in a couple of ways one of which can be related to what (Rust, 1994) calls a Newton step in the three step estimation procedure when all the model parameters that may be pre-estimated in the first steps are re-estimated altogether. Coefficients vectors of the estimated

---

[cc] Refers to the gender of the main spouse, see section 3.1.

[dd] Major reason for this outcome is the lack of relevant explanatory variables – even the age of the secondary spouse is not available in the state vector. As it was many times mentioned, expanding the state vector is very costly computationally.

regressions can similarly be included into the set of model parameters and corrected in the general maximization of likelihood function. Their large number however makes this approach very hard to apply in my case.

Therefore I introduce the following procedure which seems sensible from the computational point of view but allows for great improvement in the accuracy of the utility calculation. Instead of using the point estimates for the different sources of income, the estimated standard deviation may be incorporated into the utility calculation in the following manner. Under standard OLS assumptions the error terms are independent identically normally distributed random variables with given standard deviations. Noting that normal distribution is stable with respect to summation and that the estimated sources of income only appear in the utility function as a sum, it becomes possible to calculate the expected value of the current utility with respect to the noise in the income predictions. If the household income is represented by $I(t,d_t,s_t)=\sum_{k\in K(d_t,s_t)} I_k(t,d_t,s_t)$ where $I_k(t,d_t,s_t)=g_k(t,d_t,s_t)+\varepsilon_k$ are the estimated equations for income and the set $K(d_t,s_t)=K(sp_t,ps_{t+1})\subset\{1,2,3,4\}$ contains the indexes of the income sources relevant for the given state, the following formula of the utility can be applied.

$$
\begin{aligned}
u(d_t,s_t) &= \frac{a}{\lambda}E_\varepsilon\left\{\left(Tx(\sum_{k\in K(d_t,s_t)} I_k(t,d_t,s_t))\right)^\lambda -1\right\}+b(\overline{s}_t)\cdot L(t,d_t,s_t)+\sum_{k=0}^{7}c_k\cdot\xi\left(ps_{t+1}=k\right)\\
&=\frac{a}{\lambda}\int_{-\infty}^{+\infty}\left\{\left(Tx(\sum_{k\in K(d_t,s_t)} I_k(t,d_t,s_t))\right)^\lambda -1\right\}dF_\varepsilon+b(\overline{s}_t)\cdot L(t,d_t,s_t)+\sum_{k=0}^{7}c_k\cdot\xi\left(ps_{t+1}=k\right)\\
&=\frac{a}{\lambda}\int_0^1\left\{\left(Tx\left(\sum_{k\in K(d_t,s_t)} g_k(t,d_t,s_t)+\sqrt{\sum_{k\in K(d_t,s_t)}\sigma_k^2}\cdot\Phi^{-1}(\tau)\right)\right)^\lambda -1\right\}d\tau+\\
&\quad+b(\overline{s}_t)\cdot L(t,d_t,s_t)+\sum_{k=0}^{7}c_k\cdot\xi\left(ps_{t+1}=k\right),
\end{aligned}
\qquad (37)
$$

where $\varepsilon=\sum_{k\in K(d_t,s_t)}\varepsilon_k$ retains normal distribution with zero mean and standard deviation $\sqrt{\sum_{k\in K(d_t,s_t)}\sigma_k^2}$ approximated from the individual estimated standard errors of residuals σ_k and $\Phi(\bullet)$ is standard normal cdf. Expression (37) can be evaluated very fast with reasonable accuracy using well known Gaussian quadrature, thus it does not present additional computational burden. Recalling, that functions $g_k(t,d_t,s_t)$ are presented in Table 21-Table 25 in the Appendix, the deterministic part of the utility function is thus fully defined.

This completes the definition of preferences introducing a vector $(a, b(\overline{s}_t), c_0, ..., c_7, \lambda)$ of new parameters into the likelihood function. Thus, I consider the coefficient estimates in the income predicting models as constants integrated into the utility function. This approach may limit the overall fit of the model but does not interfere with the desired asymptotical properties of the maximum likelihood estimates of the rest of the parameters.

4.2. Motion rules (cont.)

This section concludes the definition of the conditional transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ governing the Markov stochastic process $\{\tilde{d}_t, \tilde{s}_t\}_\delta$ induced by a decision rule δ. I started defining the structure of this matrix in section 2.5 with assumption of independence of Markov health process $\{\tilde{h}_t\}$ defined with the transition probability matrix (15) and conditional on health Markov matching process $\{\tilde{m}_t\}$ defined with the transition probability matrix (16). Besides, the deterministic evaluation of the variable ps_t was described there with the means of Table 1, and the data analysis part presented essential information to define time dependent and gender specific transition probability matrices

$$\{\pi_{tij}^{(sp)}\}_{i,j\in\{0,1\}} = \begin{bmatrix} 1.0 & 0.0 \\ \pi_{t01}^{(sp)} & 1-\pi_{t01}^{(sp)} \end{bmatrix}, \tag{38}$$

which govern evaluation of the Markov process $\{s\tilde{p}_t\}$ indicating the existence of a spouse. Here $1-\pi_{t01}^{(sp)}$ represents the sample specific probability of family survival from period *t*-1 to period *t* controlled for the gender of the household as defined in the previous section.

It is now left to define the motion rules for the state variables e_t (AFP eligibility), nw_t (number of consecutive years with sufficient incomes) and aw_t (aggregate wage calculated as the average of best 20 annual earnings up to a given age). I first present the marginal motion rules for these variables (as if their were completely independent) and then return to discussion of their dependence pattern.

According to the assumption made in section 2.5 the AFP eligibility indicator e_t is deterministically defined (equalized to zero) outside the time interval between the individual specific AFP retirement age and the normal retirement age of 67. Thus, the motion rule for e_t only determines its evaluation within this time interval and can naturally be divided into two states: initial AFP eligibility forecasting and later AFP forecasting. The essential difference between these two stages is the existence of the value e_{t-1} of the AFP eligibility in the previous period.

A natural way to model probabilities of becoming or staying AFP eligible is to estimate a logit model with dichotomous outcome. However, coefficients of such model will directly enter the parameter vector of the likelihood function as the parameters of the transition probabilities matrix. Thus, care should be expressed in picking the covariates because the goodness of fit and explanatory power are only important up to the general tractability of the model. One way of reducing the number of parameters in the logit estimation is utilizing deterministic relationships between AFP eligibility and other state variables which would allow for perfect predictions. In the unidentified groups (preferably small) randomizing the motion of e_t could be done in a very "parameter efficient" way.

AFP eligibility rules described in section 3.4 give a perfect example of such deterministic relationships embedded into the data. Indeed, employment in the AFP company (expressed as $ps_{t+1} \geq 6$) and having been earning substantial wage in the last two years ($nw_t \geq 2$) can be directly checked with the state variables. Furthermore, at the ages between AFP age and normal retirement age, previously recorded AFP eligibility gives perfect prediction of success ($e_{t-1} = 1 \Rightarrow e_t = 1$), which only leaves 20 263 "undecided" individuals out of 415 124 on the second stage (only 4.88%). The outcome for this group is modeled with two coefficient logit model which simply predicts gender specific probabilities.

In the initial prediction in the absence of previous period value things are a little less perfect. It is possible to reformulate another AFP eligibility rule that puts a limitation on the average of the best 10 annual earnings in terms of the variable aw_t (representing the average of the best 20 annual earnings), but there is no other dimension in the s_t vector along which AFP eligibility perfectly separates. Data analysis shows that condition $aw_t < 74$ – the level that roughly corresponds to 2G in 1992 prices – perfectly predicts failures. One more AFP rule requires at least 10 years with substantial income after age of 50 – this may be approximated by nw_t but with no deterministic relation since the rule does not require consecutive employment. In total, 96 704 out of 179 215 (53.96%) individuals fall into the perfectly predicted group in the initial AFP forecasting and the rest are assigned initial AFP eligibility as predicted by the model presented in the Table 5 along with the one just described above.

Even though the goodness of fit for these models is rather poor for the remaining group of individuals, the fact that most of the predictions are actually perfect outside the model makes it good enough to guess AFP eligibility just a little bit better that at random. Thus, transition probability matrix for AFP eligibility $\{\pi_{ij}^{(e)}(s_t)\}_{i,j\in\{0,1\}}$ is wholly calculated on the bases of Table 5.

Table 5. Logit models of individual AFP eligibility at period t.

		Dependent variable	AFP eligibility					
		Model	Logit			Logit		
Estimated on (filter)		Age	*afp*			*afp*-67		
	Verifiable AFP rules		$ps_{t+1} \geq 6$			$ps_{t+1} \geq 6$		
			$nw_t \geq 2$			$nw_t \geq 2$		
		Other	$aw_t > 74$			$e_{t-1} = 0$		
Perfect predictions ($\Rightarrow$)			$ps_t < 6 \Rightarrow e_t = 0$ $nw_t < 2 \Rightarrow e_t = 0$ $aw_t < 74 \Rightarrow e_t = 0$			$ps_t < 6 \Rightarrow e_t = 0$ $nw_t < 2 \Rightarrow e_t = 0$ $e_{t-1} = 1 \Rightarrow e_t = 1$		
Number of perfectly predicted observations			96 704			394 861		
Applied to (filter)		Age	*afp*			*afp*-67		
	Verifiable AFP rules		$ps_{t+1} \geq 6$			$ps_{t+1} \geq 6$		
			$nw_t \geq 2$			$nw_t \geq 2$		
		Other	$aw_t > 74$			$e_{t-1} = 0$		
Variable			coef.	estim.	st.err.	coef.	estim.	st.err.
Constant term			$c_1^{(e)}$	-0,119	0,044	$c_4^{(e)}$	-0,380	0,020
Gender						$c_5^{(e)}$	-0,713	0,031
9 consecutive years with high wage $nw_t = 9$			$c_2^{(e)}$	1,165	0,042			
10 consecutive years with high wage $nw_t = 10$			$c_3^{(e)}$	2,190	0,104			
Number of observations			82 511			20 263		
Pseudo R-square (McFadden's ratio)			3,54 %			2,16%		

The motion rule for variable nw_t needs a little introduction. As shown in section 3 the number of consecutive years with high earnings moves in a particular period either upward by 1 (until it reaches the ceiling of 10) or down to 0, that is evaluates according to the following formula:

$$nw_t = \begin{cases} \min(10, nw_{t-1} + 1) \text{ with } \Pr\{w_{t-1} \geq 1G\}, \\ 0 \text{ with } \Pr\{w_{t-1} < 1G\}, \end{cases} \tag{39}$$

where w_{t-1} denotes the previous period employment income. This implies that the motion rule for nw_t can as well be described by a logit model with dichotomous dependent variable indicating either a step up or a reset down to zero. In general it could be possible within the model to introduce the dependence of this motion on the forecasted in the previous period wage, but there are two reasons not to do so. First, assigning separate parameters to these related processes gives the model more flexibility. Possible restriction could be then tested to confirm or reject the model structural assumptions. Second, as it will be clear from the next two sections, separability of transition and preferences parameters is highly desirable in structural dynamic programming models as such separability greatly simplifies the estimation process.

Since the number of consecutive years with high income is crucially dependent on previous period income, it would be natural to use the same explanatory variables as in the wage equation. Indeed, the logit estimated on the constant, previous value nw_{t-1}, aggregate wage and dummies for the previous period labour market state results in a very good fit with the McFadden's ratio[ee] 75.28%. A disadvantage of this model is the number of parameters that have to be re-estimated in the likelihood maximization. The complication is however solvable as it appears that the information carried out by these covariates can be extracted with a small number of "sufficient statistics": last period value nw_{t-1} is mainly needed to indicate $nw_{t-1}=0$, aggregate wage does not contain more information than the previous labour market state dummies, and the latter can be combined into three categories. Final model for the motion of nw_t is given in Table 6.

Table 6. Motion rule for the number of consecutive years with sufficient wage income.

Dependent variable	Increase in the number of consecutive years with high wage income $nw_t=\max\{nw_{t-1},10\}$		
Model	Logit		
Estimated on (filter)	all		
Applied to (filter)	all		
Variable	coef.	estim.	st.err.
No consecutive high income years previously $nw_{t-1}=0$	$c_1^{(nw)}$	-4,444	0,009
Active on labour market in the previous period $ps_t \in \{3,4,6\}$	$c_2^{(nw)}$	4,057	0,010
Partial disability in the previous period $ps_t \in \{5,7\}$	$c_3^{(nw)}$	2,327	0,012
OLM or fully disabled in the previous period $ps_t \in \{0,2\}$	$c_4^{(nw)}$	-5,591	0,081
Constant term	$c_5^{(nw)}$	-0,422	0,009
Number of observations	2 268 837		
Pseudo R-square (McFadden's likelihood ration)	75,23 %		

Table 7. Motion rule for aggregate wage.

Dependent variable	Aggregate wage in the current period		
Estimated on (filter)	all		
Applied to (filter)	all		
Variable	coef.	estim.	st.err.
Aggregate wage in the previous period	$c_1^{(aw)}$	1,0002	2,0E-5
Active on labour market in the previous period $ps_t \in \{3,4,6\}$	$c_2^{(aw)}$	2,6950	0,004
Constant term	$c_3^{(aw)}$	0,1440	0,005
Number of observations	2 268 837		
Pseudo R-square (McFadden's likelihood ration)	99,91 %		
Estimated standard error of residuals	2,6585		

[ee] More on this type of measure in section 6.5.

It is now only left to describe the motion rule for the aggregate wage aw_t. This is done with a simple regression model presented in Table 7. Aggregate wage moves in a very smooth way almost completely explained by its previous value (See Figure 24 in the Appendix, p.107). Still, the variable indicating active labour market states turns out to be very significant (with the t-value of 678.18) and helps to distinguish the cases when aggregate wage actually grows from those when it is practically unchanged (first of all, pension). Positive coefficients of the regression ensure non-decreasing behavior implied by the definition. Extremely high R-square and relatively small estimated standard error of residuals allows me to treat this relationship as deterministic which allows for significant simplification of the calculation procedures.

Once all state variables are assigned their own motion rules, I can discuss the implied pattern of dependence in the transition matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$. First, as assumed earlier, the health process $\{\tilde{h}_t\}$ is completely independent in a sense that it evaluates only conditional on its own previous value. The same property is true for spouse existence process $\{s\tilde{p}_t\}$ which appears completely deterministic and exogenous (although individual specific). The matching process dependent only on health in the same period combined with health process form an independent pair $\{\tilde{h}_t, \tilde{m}_t\}$. Previous labour market state process $\{ps_t\}$ is deterministic but depends directly on own previous value and the decision in the previous period but also indirectly through current health, match and AFP eligibility on their previous period values. As specified by the logit model in Table 6 $\{n\tilde{w}_t\}$ process depends on own previous values as well as previous labour market state and thus indirectly on previous health, match and AFP eligibility. Aggregate wage process $\{aw_t\}$ is dependent on the previous labour market state besides own previous value. Finally, AFP eligibility process $\{\tilde{e}_t\}$ has the most complicated dependence pattern since the logit models in Table 5 describing its movements are dependent on previous period own value, number of consecutive sufficient annual wages and the aggregate wage, but also through the previous labour market state on previous health and match. All this is not to mention dependence on age, gender and AFP eligibility age.

The described dependence pattern is graphically displayed in Figure 1. Solid lines represent direct dependence of period *t*+1 values on period *t* values whereas dashed lines represent indirect dependence links which go from indicated variable in *t* to the indicated variable in *t*+1 through some other variable in *t*+1. Dependence on time, AFP age and gender of the household are not shown in the picture.

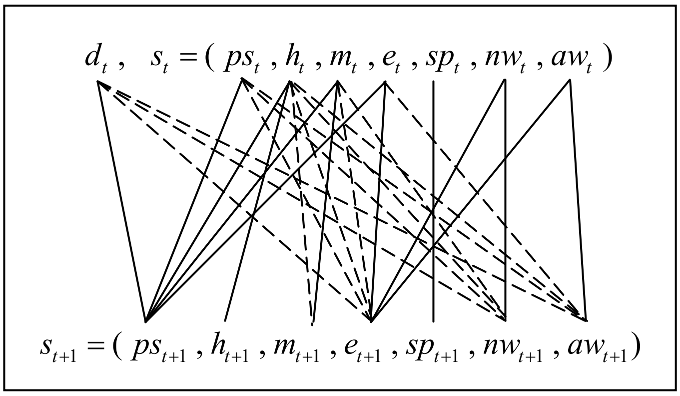

Figure 1. Dependence pattern in the transition probability matrix.

The constructed pattern of dependence in the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ appears to be quite complex and requires special attention when implemented in the computational algorithm. Table 26 in the Appendix (p. 99) displays the implementation of the function that returns transition probability for given values of current and next period state variables. The function deals with the total transition probability as a chain starting from the independent health transition probability and then multiplied by conditional probabilities related to each of the state variables which are carefully calculated taking into account all dependencies presented in Figure 1. The function constructs well defined probability distributions for each allowed combination of the current state and decision variables as defined by the assumptions listed in section 2.5 (and corresponding to the black areas in Figure 11 in the Appendix).

This concludes the definition of the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$ giving final appearance to the dynamic model apart from two minor functions used in preferences and defined in the next two short sections. Complete definition of transition probabilities introduces additional the parameter vectors $c^{(e)} = (c_1^{(e)},..,c_5^{(e)})$, $c^{(nw)} = (c_1^{(nw)},..,c_5^{(nw)})$ for the likelihood function.

4.3. Tax function

The tax function $Tx(\bullet)$ is a mapping from the set of before tax household incomes into the set of the disposable household incomes which is superposed inside the utility function. In order not to complicate the construction of the dynamic model, the Norwegian tax system has to be represented within the scope of the model. This section presents the technique used in this respect.

The fully detailed Norwegian tax function is available due to (Haugen, 2000). Main determinants of the household taxes are: year of calculation, type of household (single or full), distribution of the household income into employment income, capital income and pensions, distribution of household income among spouses, geographical location of the household and other information. Clearly, most of these characteristics are not available within the model. Again, a statistical approach can be applied to relate the accurately calculated taxes to the state variables available in the model.

To reconstruct the needed statistical relationship I use simulated data build on the following principles. First, the data contains all combinations of the following variables in ever imaginary year form 1992 to 2001 (altogether 160 160 data points):

- Household before tax income between 1 and 1000 measured in 1000 NOK in 1992 prices.
- Spouse indicator sp_t (0 or 1).
- Current labour market state recorded in ps_{t+1} (8 values).

Second, the following assumptions about the household income distributions are made. If $sp_t = 0$ the whole income is generated by the primary spouse, and depending on ps_{t+1} variable is assumed to originate fully from social security sources (when $ps_{t+1} \in \{0,1,2\}$), fully from employment sources (when $ps_{t+1} \in \{4,6\}$, or to be equally divided between the two (when $ps_{t+1} \in \{3,5,7\}$). If $sp_t = 1$ one third of household income is assigned to the spouse who is assumed to be a pensioner as a result of a simple random draw with probability $\frac{1}{2}$. In this case the part of the household income corresponding to the spouse is categorized as pension, otherwise it is categorized as wage. The two thirds of household income corresponding to the primary spouse are assigned a source the same way as in single households.

Third, the full detail tax function is applied to the simulated data calculating taxes in every of the 160 160 cases. Calculated taxes are then used in the simple regression model which is estimated with the results presented in Table 27 in the Appendix (p. 100).

As follows from the regression results, the model displays very good fit (R-square 98.34%) along with very precise coefficients. The marginal tax for household incomes is estimated at the level 42.3% which is rather high, but compensated by the negative coefficients for all the explanatory variables most extensively for full households with primary spouse on pension or disability. High measure of goodness of fit allows treating the estimated tax function as deterministic, also keeping in mind that potential inaccuracies do not compromise the asymptotic properties of maximum likelihood estimation, but only affect the strict definition of the utility function.

4.4. Termination function

The termination value function represents the residual utility after the age of 70 when the process $\left\{\tilde{d}_t, \tilde{s}_t\right\}_\delta$ can no more be controlled by the decision variable. In general following (1) it can be written as the infinite sum of the uncontrolled utilities multiplied by time specific depreciation rate:

$$\Lambda(s_T) \approx \sum_{t=T+1}^{\infty} \rho_t \beta^{t-T_0} U(\tilde{s}_t). \tag{40}$$

In order to represent (40) in terms of the existing state variables assume their motion freezes at termination period T and the utility after T becomes constant. If furthermore survival probabilities ρ_t are independent of time beyond T, (40) becomes a geometric series and the termination function can be represented by its sum

$$\Lambda(s_T) \approx \frac{\rho_T \beta^{T-T_0+1} U(s_T)}{1-\beta}. \tag{41}$$

Further, the only source of heterogeneity in the utility level after T is the last observed aggregate wage aw_T (in addition to the "constant" variables indicating household gender and the individual specific AFP age). Therefore, it is possible to abstract from the detailed utility function and represent (41) as a simple linear function of the last aggregate wage.

$$\Lambda(s_T) = c_1^{(tf)} \cdot afp + c_2^{(tf)} \cdot gender + c_3^{(tf)} \cdot aw_T. \tag{42}$$

The possible constant term in (42) is omitted because as the termination function enters the utility and correspondingly the value function only in the termination period, it would simply introduce an additively separable parameter in both enumerator and denominator in (26) and would clearly be unidentified.

The last expression completes in all details the specification of the structural dynamic model which main components are: the agent sequential decision problem (2) in which the state vector (8) represents the evaluation of the stochastic process $\left\{\tilde{d}_t, \tilde{s}_t\right\}_\delta$; the latter is induced by a feasible decision rule δ which belongs to the class $\mathfrak{F}$ and in every time period defines the optimal decision from the choice set (9) and thus affects the transition of process by the means of the matrix $\left\{p(s_t \mid s_{t-1}, d_{t-1})\right\}$ defined by (15), (16), (38) and (39). The model rationalizes the panel of observations $\left\{d_t^a, ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\right\}_{t\in\{T_0^a-1,..,T^a\}, a\in\{1,..,A\}}$ with the means of likelihood function (33) dependent on the vector of parameters

$$\begin{aligned}\theta = (&\pi_{00}^{(h)}, \pi_{01}^{(h)}, \pi_{00}^{(m)}, \pi_{01}^{(m)}, \pi_{10}^{(m)}, \pi_{11}^{(m)}, \pi_{20}^{(m)}, \pi_{21}^{(m)}, \theta^{bhm},\\ &\beta, \lambda, a, b(\overline{s}_t), c_0, \ldots, c_7,\\ &c_1^{(e)}, .., c_5^{(e)}, c_1^{(nw)}, .., c_5^{(nw)}, c_1^{(tf)}, .., c_3^{(tf)}),\end{aligned} \tag{43}$$

altogether at least 34 parameters from which 19 are associated with the transition probability matrix and at least 15 are associated with preference calculations (from which at least 11 enter instantaneous utility function, 1 reflects the time preferences and 3 enter the termination function).

5. Calibration of the model

This section presents a preliminary analysis of the behavior of the model mainly aimed at the calibration of the latent state processes. In particular I test a possible simplifying assumption on the matching process transition probability matrix that could allow for a reduction in the number of parameters.

5.1. Calibration technique

The main purpose for the preliminary calibration of the model is making the main optimization of the likelihood function easier. This can be achieved by exploring its area of definition[ff], scaling the parameters to the magnitude of 0.1-1 and finding good starting values for them. I mostly concentrate on the latent health and job match processes investigating responses to changes in their motion rules.

To simplify calibration, itself instead of computationally demanding likelihood function, I use the much simpler aggregated squared error function which measures the Euclidean distance between the observed distributions among labour market states in all the periods and its predicted counterpart. In other words, if the observed distributions are given by the matrix

$$lmd^{(obs)} = \begin{bmatrix} lmd^{(obs)}_{50,0} & lmd^{(obs)}_{51,0} & \cdots & lmd^{(obs)}_{70,0} \\ lmd^{(obs)}_{50,1} & lmd^{(obs)}_{51,1} & & lmd^{(obs)}_{70,1} \\ \vdots & & \ddots & \\ lmd^{(obs)}_{50,7} & lmd^{(obs)}_{51,7} & & lmd^{(obs)}_{70,7} \end{bmatrix}, lmd^{(obs)}_{t,k} = \frac{\sum_{a=1}^{A} \xi(ps_{t+1} = k)}{\sum_{k=0}^{7}\sum_{a=1}^{A} \xi(ps_{t+1} = k)}, \tag{44}$$

where $\xi(\bullet)$ is again an indicator function[gg], $\sum_{k=0}^{7} lmd^{(obs)}_{t,k} = 1$; and $lmd^{(sim)}$ denotes the similarly defined matrix on the simulated data, the calibration function can be specified as

$$\Delta(\theta) = \sum_{t=T_0}^{T}\sum_{k=0}^{7}\left(lmd^{(obs)}_{t,k} - lmd^{(sim)}_{t,k}\right)^2 . \tag{45}$$

Simulated data was constructed of 1000 hypothetical households which were assigned initial characteristics from the first observations of the randomly chosen sub-sample of household. The behavior of these households was forecasted as solution of the agents problem (2). Random assignment of the initial characteristics was held the same throughout the whole calibration process while the parameter vector θ was adjusted to minimize the value of the objective function (45).

[ff] Dependent on the values of the parameters and certain observed variables numerical errors as overflows and logarithms of zeros may appear while calculating the likelihood function.

[gg] Indicator function also returns zero in the cases when no observation is available.

The calibration was focused on the parameters of the latent processes of health and matching. The rest of the parameters were given the following values.

- Discount factor $\beta = 0.95$.
- Main preference parameters $\lambda = 0.5, a = 0.001, b(\overline{s}_t) \equiv 0.001, c_0 = ... = c_7 = 0$.
- Motion parameters $c^{(e)}, c^{(nw)}$ for number of consecutive years with sufficient annual wage income and individual AFP eligibility as estimated by the corresponding models (See Table 5, Table 6).
- Termination functions parameters $c_1^{(tf)} = .. = c_3^{(tf)} = 0$.

In addition transition probability matrix for matching process was assumed to have the following semi-symmetric structure which allowed for reduction of the corresponding number of parameters from 6 to 3.

$$\begin{bmatrix} \pi_{00}^{(m)} & \pi_{01}^{(m)} & \pi_{02}^{(m)} \\ \pi_{10}^{(m)} & \pi_{11}^{(m)} & \pi_{12}^{(m)} \\ \pi_{20}^{(m)} & \pi_{21}^{(m)} & \pi_{22}^{(m)} \end{bmatrix} = \begin{bmatrix} \pi_{00}^{(m)} & \frac{1}{2}(1-\pi_{00}^{(m)}) & \frac{1}{2}(1-\pi_{00}^{(m)}) \\ \pi_{10}^{(m)} & \pi_{11}^{(m)} & 1-\pi_{10}^{(m)}-\pi_{11}^{(m)} \\ \pi_{10}^{(m)} & 1-\pi_{10}^{(m)}-\pi_{11}^{(m)} & \pi_{11}^{(m)} \end{bmatrix}. \tag{46}$$

The chosen objective function (45) is about twice faster to calculate than the likelihood function and has the very convenient interpretation as the aggregated accuracy of predictions of labour market outcomes, but mathematically displays challenging properties. As seen in the Figure 35 in the Appendix (p. 113) which plots this function with respect to $\pi_{00}^{(h)}$, it is very unsmooth and partially flat. This makes calibration quite difficult and justifies the use of derivative free methods as Nelder-Mead method for its minimization.

5.2. Calibration results

Calibration was carried out in the Matlab programming interface using the built in optimization machinery. The following parameters were found by the optimization procedure and gave a decent distribution of predicted labour market states.

$$\pi^{(h)} = \begin{bmatrix} \pi_{00}^{(h)} & \pi_{01}^{(h)} & 1-\pi_{00}^{(h)}-\pi_{01}^{(h)} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.9589 & 0,0160 & 0,0251 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\pi^{(m)} = \begin{bmatrix} \pi_{00}^{(m)} & \pi_{01}^{(m)} & 1-\pi_{00}^{(m)}-\pi_{01}^{(m)} \\ \pi_{10}^{(m)} & \pi_{11}^{(m)} & 1-\pi_{10}^{(m)}-\pi_{11}^{(m)} \\ \pi_{20}^{(m)} & \pi_{21}^{(m)} & 1-\pi_{20}^{(m)}-\pi_{21}^{(m)} \end{bmatrix} = \begin{bmatrix} .0499 & .47505 & .47505 \\ .0106 & .9007 & .0887 \\ .0106 & .0887 & .9007 \end{bmatrix}, \tag{47}$$

$\theta^{bhm} = 0.8045$.

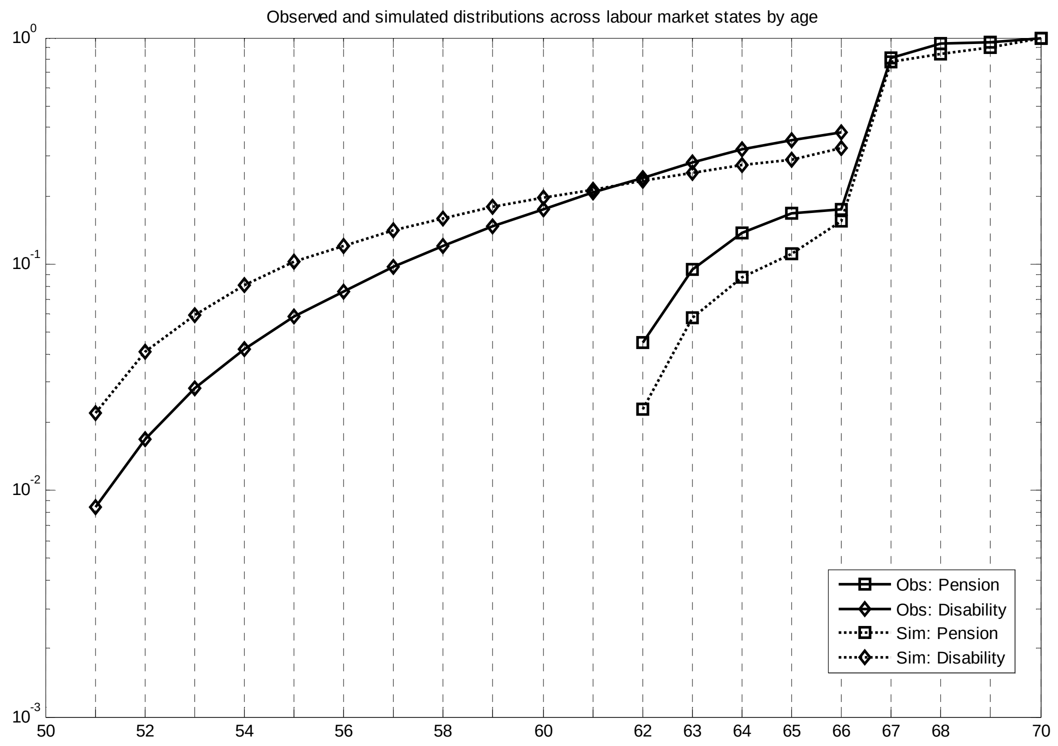

Figure 2. Observed and predicted fractions of pension and disability states.

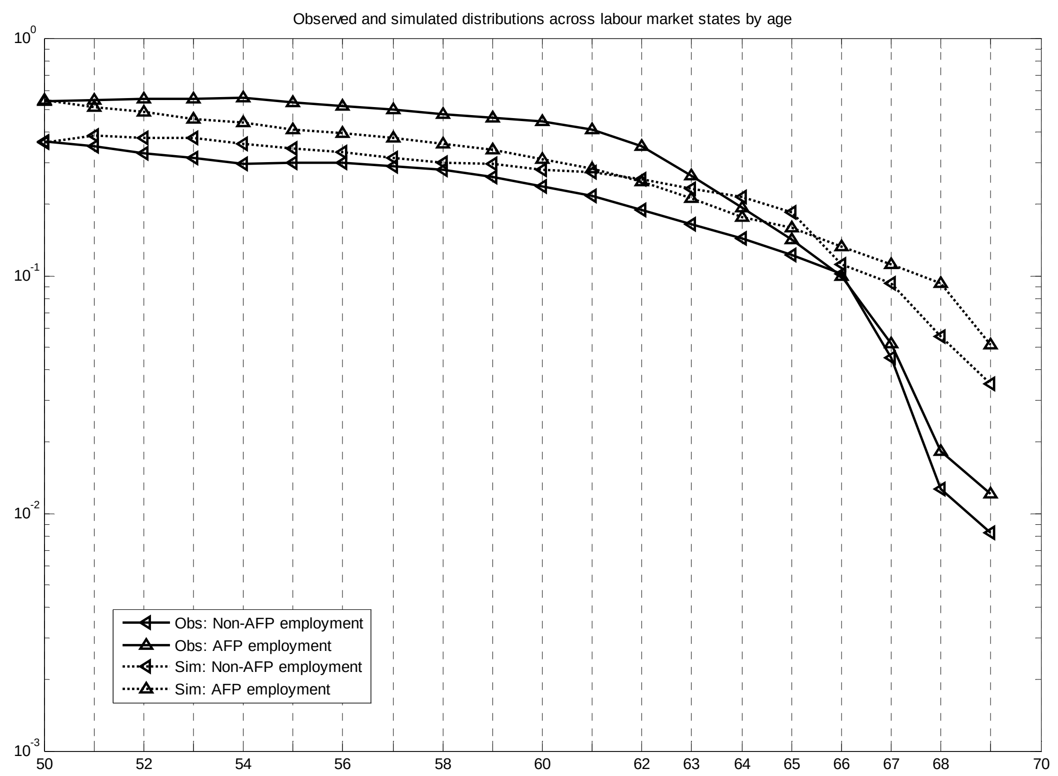

Figure 3. Observed and predicted fractions of working states.

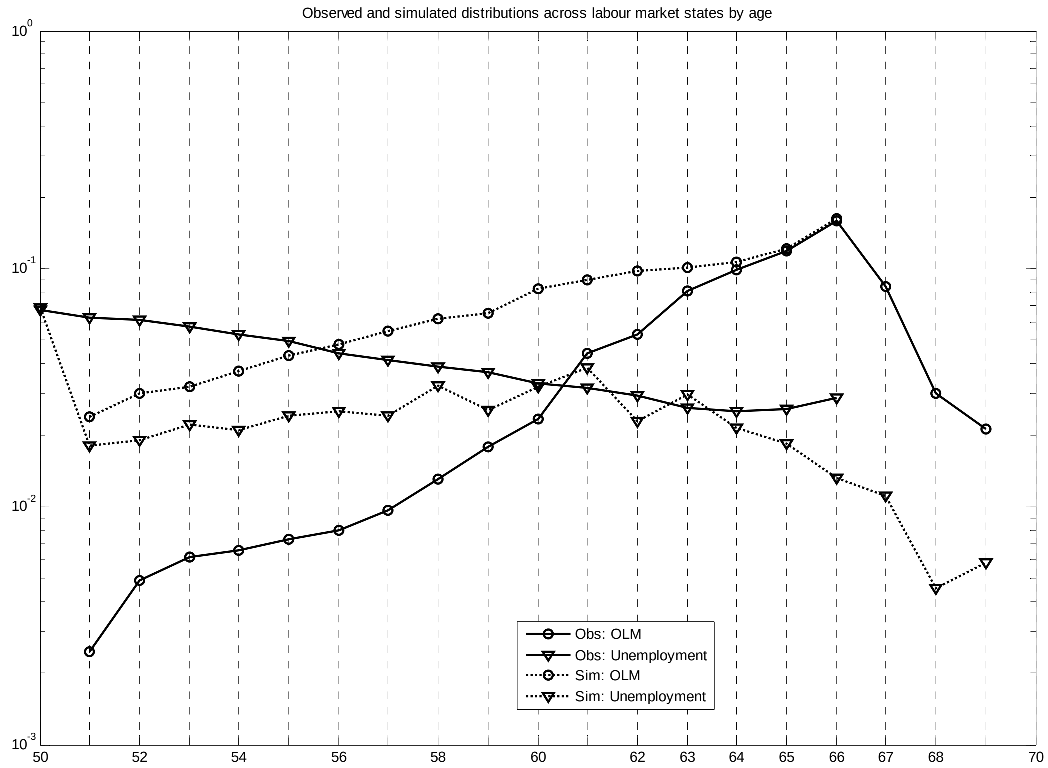

Figure 4. Observed and predicted fractions of OLM and unemployment states.

Calibration results are presented in a form of graphs that compare observed and forecasted distributions of the sample among the labour market states in Figure 2–Figure 4 above and also in the Table 14 and Table 28 in the Appendix (p. 84).

As follows from the graphs and the tables the calibrated model represents the observed behavior rather well. The most accurate predictions seem to be concentrated around ages 60-67 where the most important transitions take place and most of the observations are available. In general pension and disability states seem to be predicted best of all. Employment in AFP and non-AFP companies is correspondingly under- and overestimated, besides both types of employment are overestimated after the usual retirement age. The fact that the predictions of the two types of employment tend towards each other and fail to represent the observed in the real data gap clearly indicates severe limitation imposed by the semi-symmetric structure of the matching transition probability matrix. The kinky predictions of the unemployment state may have been affected by the relatively small number of simulations compared to the number of observed households. The underestimated pension take-up together with poor estimation of the residual labour market state shares may have been influenced by the exclusion of the preferences from calibration.

Nevertheless, calibration of the model produced reliable starting values for the parameters to be used in the full scale estimation undertaken in the next section.

6. *Estimation of the model*

This section describes model estimation methods and estimation strategy, derives final definition for the parameter vector (43), presents the estimates for the parameters and discusses the achieved goodness of fit of the model.

6.1. Estimation strategy

Maximization of the likelihood function (33) with respect to the 34 parameters listed in the parameter vector (43) presents a demanding computational task. Straightforward maximization with respect to all the parameters is especially undesirable because it may take enormous amounts of iterations before convergence. Step by step optimization with respect to different parameter groups is more appropriate. First, as it appears, vector θ contain highly interconnected parameters as discount factor β, the income coefficient a and the Box-Cox parameter λ. Even a slightest change in one of them causes large changes in the partial derivative of the likelihood function with respect to another. Second, and more challenging, a single computation of the likelihood function is very time consuming.

In the similar circumstances (Rust, 1994) suggests separating the likelihood function (28) into two partial likelihood functions composed correspondingly of the transition probabilities $p(s_t^a \mid s_{t-1}^a, d_{t-1}^a, \theta)$ and choice probabilities $P_t(d_t^a \mid s_t^a, \theta)$ under the assumption of separability of the parameter vector. The resulting combined vector of the estimates is consistent and asymptotically normal but lacks efficiency. To improve this estimate further Rust suggests an additional third step which is virtually one iteration of a quasi-Newton line search algorithm, and shows that three-step procedure gives the result which is asymptotically equivalent to the full information maximum likelihood.

This decomposition of the problem gives a great reduction of the computational burden of the likelihood function maximization. In the current model the parameter vector is indeed composed of the two subvectors corresponding respectively to preferences and transitions thus granting the required separability:

- 19 parameters associated with the transition probability matrix $\{p(s_t \mid s_{t-1}, d_{t-1})\}$, namely the health transition probabilities, the matching transition probabilities along with the reduction parameter θ^{bhm}, and the parameters of the logit models for AFP eligibility and the number of consecutive high annual earnings.

- at least 15 parameters associated with instantaneous and intertemporal preferences, namely the discount factor β, the parameters of the utility function and the parameters of the termination function.

I adopt the following four step estimation strategy (as illustrated by Table 8), which is inspired by (Rust, 1994) and builds on the general two-step maximum likelihood procedure described, for example in (Greene, 2000). In the first step all the separable parameters are estimated with the independent models. This step was already performed in section 4.2 when relevant data was used to fit logit models for e_t and nw_t motion rules[hh]. Altogether, it covers 10 parameters and the estimates are reported in Table 5 and Table 6 in section 4.2 (pp. 53-54).

Table 8. Estimation strategy.

Parameters	Number of parameters	Description	Estimated in the steps			
			1	2	3	4
$\pi_{00}^{(h)}, \pi_{01}^{(h)}$	2	Health transition probability matrix		■		■
$\pi_{00}^{(m)}, \pi_{01}^{(m)}, \pi_{10}^{(m)}, \pi_{11}^{(m)}, \pi_{20}^{(m)}, \pi_{21}^{(m)}, \theta^{bhm}$	7	Matching transition probability matrix		■		■
$c_1^{(e)},..,c_5^{(e)}$	5	AFP eligibility indicator motion rule	■	■		■
$c_1^{(nw)},..,c_5^{(nw)}$	5	Number of consecutive high annual wage incomes motion rule	■	■		■
β	1	Discount factor			■	■
$\lambda, a, b(\overline{s}_t), c_0,...,c_7$	≥11	Utility function parameters			■	■
$c_1^{(tf)},..,c_3^{(tf)}$	3	Termination function parameters			■	■

The second step estimates the parameters of the transition probability matrix which are separable from preferences (19 parameters). The approach proposed by (Rust, 1994) must be somewhat modified here because of the additional complication in the likelihood function (33) caused by the latent state variables. To maintain on average the random structure imposed by the choice probabilities I replace them with the uniform distribution instead of completely dropping them from the likelihood function (which is equivalent to replacing them with ones). This is important because the choice probabilities dependable on the choice sets implied by the latent variables influence the identification of the transition probability parameters. In other words, in the second estimation step I use the following partial likelihood function:

[hh] All the models were estimated either by maximum likelihood or by OLS under classical assumptions which makes the estimates asymptotically equivalent to those of maximum likelihood estimation.

$$L_1(\theta') = \prod_{a=1}^{A} \left(\sum_{(h,m)^a \in HM^a} p_0(m_{T_0^a}, h_{T_0^a}, \theta') \prod_{t=T_0^a+1}^{T^a} \frac{1}{\left|D_t(s_t^a)\right|} \cdot p(s_t^a \mid s_{t-1}^a, d_{t-1}^a, \theta') \right),$$
$$s_t^a = \left(ps_t^a, h_t, m_t, e_t^a, sp_t^a, nw_t^a, aw_t^a \right), \tag{48}$$

where $\left|D_t(s_t^a)\right|$ denotes the number of alternatives in the choice set $D_t(s_t^a)$ available for an agent at time t and $\theta' = (\pi_{00}^{(h)}, \pi_{01}^{(h)}, \pi_{00}^{(m)}, \pi_{01}^{(m)}, \pi_{10}^{(m)}, \pi_{11}^{(m)}, \pi_{20}^{(m)}, \pi_{21}^{(m)}, \theta^{bhm}, c_1^{(e)}, .., c_5^{(e)}, c_1^{(nw)}, .., c_5^{(nw)})$. Note that the parameters of the logit models $(c_1^{(e)}, .., c_5^{(e)}, c_1^{(nw)}, .., c_5^{(nw)})$ are re-estimated in this step. The rest of the parameters are taken as given forming a two-stage structure. Greene (2000) references (Murphy and Topel, 1985) who show that the resulting estimates retain the properties of consistency and asymptotical normality with an adjusted covariance matrix.

In the third step I define $b(\overline{s}_t)$ and estimate the parameters associated with the preferences by the means of maximization of the complete likelihood function (33) with respect to the parameters $\theta'' = (\lambda, a, b(\overline{s}_t), c_0, ..., c_7, c_1^{(tf)}, .., c_3^{(tf)})$ keeping parameter subvector θ' fixed. Again, the resulting two-stage procedure again yields consistent and asymptotically normal estimates.

To conclude, the forth step performs the final maximization of the full likelihood function with respect to all the parameters where the previous estimates are used as starting values. With the logic of (Rust, 1994) since all previously estimated parameters are consistent, just a few quasi-Newton iterations are necessary for convergence. Besides yielding asymptotically efficient estimates, performing the final step as independent maximization allows me to skip complicated calculations of the covariance matrix for the parameters, which could in principle be done based on the mentioned results of Rust and Murphy. Instead, I use "information equality" to approximate the covariance matrix from the numerically computed Hessian at the solution point of the final maximization problem.

The proposed decomposition of the maximization procedure helps to defy the main source of the computational difficulty in the problem, namely the complicated recalculation of the value function which is required at each evaluation of the likelihood function. The value function has to be calculated on each point of the state space[ii] in each time period before the choice probabilities $P_t(d_t^a \mid s_t^a, \theta)$ may be constructed according to (26). This separates the value function calculation into a distinct computational task which is independent of any manipulations with the data. Since the number of function evaluations in maximization routine grows very fast with the number of

[ii] With 7 points grid for the aggregate wage, the total number of points of the state space is 133 056.

parameters, especially when numerical derivatives are used, one way to deal with the complication is to avoid the value function calculation as much as possible. Accordingly, the chosen estimation strategy requires value function calculations only on the last two steps of the process when half of the parameters are already pre-estimated.

Another source of computational difficulty is integration over the unobserved variables. As described in section 2.6 all trajectories of the partial state process $\{\tilde{h}_t, \tilde{m}_t\}$ consistent with the observed state variables $\{ps_t^a, e_t^a, sp_t^a, nw_t^a, aw_t^a\}_{t\in\{T_0^a-1,..,T^a\}}$ for each household $a \in \{1,.., A\}$ have to be followed by the likelihood calculating routine. Since at each time period observed state vector is likely to allow for several values of health, this essentially means following the branches of a tree[jj] which in the worst case has 21 levels with up to 3 branches at each level. Unfortunately, this process is hard to optimize because the branching depends on both observations (to ensure latent variable consistency) and on the previous values of the state variables (to ensure model structure). Thus, the runtime of the part of the likelihood calculation routine which deals with the data and integrates out the unobserved health and match processes crucially depends on the number of agents in the sample. The full dataset comprises 200 921 households which provide 2 165 467 observations (see section 3.1).

The model was implemented in the MatLab environment with the inner circuit of the nested fixed point algorithm (see detailed description in (Rust, 1987)) programmed as a dynamically linked library written in C programming language and the outer circuit completely taken care of by the standard MatLab unconstrained minimization routine[kk]. Using C for the computationally demanding part of the program (both value function calculation and the integration over the unobserved variables) ensured obtaining minimal running time. On the Frisch Center server[ll] one evaluation of the likelihood function over a half of the dataset took approximately 160-170 seconds with approximately 60 seconds spent on the value function calculation. Production runs of the optimization routine were performed in the supercomputing center of the University of Oslo[mm]. Distributing the computational load over 8 to 12 CPUs brought the one evaluation run-time over the

---

[jj] The routine is implemented as a recursive function in C.

[kk] Namely, fminunc() routine performed a line search with numerical derivatives and the BFGS Hessian updating followed by at most three single Newton steps based on the numerically approximated Hessian for “fine tuning” into the optimal point.

[ll] At the time Anton: Dell PowerEdge 2850 x64-based PC with 8 EM64T Family 15 Model 4 Stepping 8 GenuineIntel ~2793 Mhz processors, 8GB physical memory, running Microsoft Windows Server 2003 Enterprise x64 Edition, version 5.2.3790 Service Pack 1 Build 3790. 32bit MatLab version 7.3.1.267 (R2006b).

[mm] The Titan II cluster at the time comprising 1852 CPUs (on 4-core SUN X2200 AMD and 2-CPU DELL 1425 Intel nodes) was used, see http://www.hpc.uio.no, http://login3.titan.uio.no/ganglia/.

full sample down to 9-12 seconds.[nn] Parallelization was applied to both value function calculation and the data processing.

Using an unconstrained optimization routine required reformulation of the parameters $(\pi_{00}^{(h)}, \pi_{01}^{(h)}, \pi_{00}^{(m)}, \pi_{01}^{(m)}, \pi_{10}^{(m)}, \pi_{11}^{(m)}, \pi_{20}^{(m)}, \pi_{21}^{(m)})$ constrained by the obvious probability limitations in the following manner.

$$\begin{aligned}
\pi_{00}^{(h)} &= \frac{\exp(p_{00}^{(h)})}{1+\exp(p_{00}^{(h)})+\exp(p_{01}^{(h)})}, \pi_{01}^{(h)} = \frac{\exp(p_{01}^{(h)})}{1+\exp(p_{00}^{(h)})+\exp(p_{01}^{(h)})}, \\
\pi_{00}^{(m)} &= \frac{\exp(p_{00}^{(m)})}{1+\exp(p_{00}^{(m)})+\exp(p_{01}^{(m)})}, \pi_{01}^{(m)} = \frac{\exp(p_{01}^{(m)})}{1+\exp(p_{00}^{(m)})+\exp(p_{01}^{(m)})}, \\
\pi_{10}^{(m)} &= \frac{\exp(p_{10}^{(m)})}{1+\exp(p_{10}^{(m)})+\exp(p_{11}^{(m)})}, \pi_{11}^{(m)} = \frac{\exp(p_{11}^{(m)})}{1+\exp(p_{10}^{(m)})+\exp(p_{11}^{(m)})}, \\
\pi_{20}^{(m)} &= \frac{\exp(p_{20}^{(m)})}{1+\exp(p_{20}^{(m)})+\exp(p_{21}^{(m)})}, \pi_{21}^{(m)} = \frac{\exp(p_{21}^{(m)})}{1+\exp(p_{20}^{(m)})+\exp(p_{21}^{(m)})}.
\end{aligned} \tag{49}$$

Parameter θ^{bhm} limited to the unit interval was modified so that

$$\theta^{bhm} = \frac{1}{1+\exp(\theta_m^{bhm})}. \tag{50}$$

Invariance of the maximum likelihood estimation and the theorem about the asymptotic distribution of the nonlinear function (Theorem 4.17 in (Greene, 2000)) provide a mechanism of backward recalculation of the estimates and their covariances according to (49) and (50). In the next section I shall skip these details and report the original parameters with their estimated standard errors directly.

Finally, using standard MatLab minimization routines required multiplying the likelihood function by (-1) which also influences the calculation of the information matrix and the covariance matrices. In addition I scaled the logarithm of the likelihood function to the order of magnitude of one to facilitate the optimization. All these adjustments are undone before the results are displayed in the next section.

6.2. Estimation results: step 2.

Whereas the results of the first step were already presented in section 4.2. in Table 5 and Table 6 this section takes the second step and presents the conditional transition probability estimates.

The calibration results presented in section 5.2 (transformed in accordance to (49) and (50)) provided good starting values for the maximization of the partial likelihood function (48). The

---

[nn] Run-time decreased non-linearly in the number of CPUs and had a pronounced minimum at 8-12 CPUs.

latter appeared to be smooth and well-behaved, so that direct maximization over all 19 corresponding parameters went with no complications and achieved convergence after 93 iterations and 4 873 function evaluations. Since partial likelihood was used, the optimal function value is not comparable with other results. Table 9 displays the estimates and their asymptotic standard errors in comparison to the calibration results (for the health and match probabilities) and previous estimates (for AFP eligibility indicator).

Table 9. Estimates of the transition probability matrix parameters at step 2.

Parameter description			Previous estimate		Current estimate		
			Estimate	Std. Err.	Estimate	Std. Err.	p-value
Health transition probabilities	Good to good health	$\pi_{00}^{(h)}$	0.9589	-	0.97073	0.00011	0.00000
	Good to bad health	$\pi_{01}^{(h)}$	0.0160	-	0.02806	0.00011	0.00000
Matching transition probabilities	No match to non-AFP	$\pi_{01}^{(m)}$	0.0499	-	0.19737	0.00151	0.00000
	No match to AFP	$\pi_{02}^{(m)}$	0.4751	-	0.13229	0.00131	0.00000
	Non-AFP to non-AFP	$\pi_{11}^{(m)}$	0.0106	-	0.89112	0.00045	0.00000
	Non-AFP to AFP	$\pi_{12}^{(m)}$	0.9007	-	0.08915	0.00042	0.00000
	AFP to non-AFP	$\pi_{21}^{(m)}$	0.0106	-	0.05584	0.00026	0.00000
	AFP to AFP	$\pi_{22}^{(m)}$	0.0887	-	0.93550	0.00027	0.00000
Matching limitation due to bad health		θ^{bhm}	0.8045	-	0.99996	0.00008	0.00000
Motion rule for e_t	Constant term (1)	$c_1^{(e)}$	-0.119	0.044	-0.31837	0.05124	5.21E-10
	$nw_t = 9$	$c_2^{(e)}$	1.165	0.042	1.24310	0.10514	2.95E-32
	$nw_t = 10$	$c_3^{(e)}$	2.190	0.104	2.36392	0.05245	0.00000
	Constant term (2)	$c_4^{(e)}$	-0.380	0.020	-0.30423	0.01949	6.21E-55
	Gender	$c_5^{(e)}$	-0.713	0.031	-0.62980	0.02928	1.31E-102
Motion rule for nw_t	$nw_{t-1} = 0$	$c_1^{(nw)}$	-4.444	0.009	-5.78499	0.02122	0.00000
	$ps_t \in \{3,4,6\}$	$c_2^{(nw)}$	4.057	0.010	5.64890	0.01415	0.00000
	$ps_t \in \{5,7\}$	$c_3^{(nw)}$	2.327	0.012	13.73441	1.44000	1.46E-21
	$ps_t \in \{0,2\}$	$c_4^{(nw)}$	-5.591	0.081	-5.93205	0.09033	0.00000
	Constant term	$c_5^{(nw)}$	-0.422	0.009	-0.47464	0.00863	0.00000

Table 9 indicates a considerable amount of updating of the parameters especially vivid in the motion rule for variable nw_t. Although the signs and roughly the quantitative relations among the estimates remain, all the coefficients moved away from zero and thus became individually more influential in determining the direction of the next step in motion of nw_t. Moreover, previous combined disability and employment labour market state became more or less sufficient condition for the increase in nw_t variable next period. These changes induced by the structure of the model are due to the fact that the complex evolution of the dynamic environment in the model introduces additional relationships overseen in the separate static preliminary estimations.

Similar shifting of the estimates away from zero is seen in the initial AFP eligibility assignment equation. Decrease in the constant term reflects the lower average probability of becoming AFP eligible at the AFP age while the opposite neutralizes it with approximately same magnitude increase of the coefficient with the condition $nw_t = 10$ maintaining the relative importance of having solid employment in the later years. As a result, the change in the estimates reflects the decreased influence of the condition $nw_t = 9$ compared to the preliminary case which is expected since the structure of the model takes on a part of the explanatory burden. This very effect is even better illustrated in the second AFP equation (for years after the AFP age) as both of the coefficients shift towards zero and thus decrease the amount of the information contained in the logit probability distribution implied by it. Again, general structure of the dynamic model imposed onto the estimation takes on some of the explanatory burden.

Transfer probabilities for the health and matching processes are also estimated very sharply, but do change quite a lot from the calibration results. This is expected, especially in light of rejection of the semi-symmetric structure of the transition probability matrix for the matching process in section 5.2. Obtained estimates imply the following transition probability matrices.

$$\pi^{(h)} = \begin{bmatrix} 0.9707 & 0.0281 & 0.0012 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \pi^{(m)} = \begin{bmatrix} 0.6703 & 0.1974 & 0.1323 \\ 0.0197 & 0.8911 & 0.0892 \\ 0.0087 & 0.0558 & 0.9355 \end{bmatrix}. \quad (51)$$

Thus, the probability of remaining in good health is 97.07% whereas the probability of a serious health shock resulting in a sure full-time disability is about 0.12%. The model is capable of predicting this parameters with no corresponding observations due to its structural nature. Health parameters are identified from two sources of information. First, since health is partially recoverable from the data as mention numerously, there is a direct source of identification. Second, and more importantly, careful reconstruction of the sets of available choices as the main element of the structure of the model allows for identification of health parameters indirectly through likelihood maximization similar to the parameters of unobserved heterogeneity. Under the imposed structure, the obtained parameters appear most reasonable (in fact, consistent and not far from asymptotically efficient) to suggest that the estimated health process is likely to generate the observed behavioral patterns. A reasonable question in this respect could be whether the direct identification along would differ much in conclusions about the latent health. Using partial recoverability it is in principle possible to calculate the rate of observed transitions from the "healthy" labour market states (0, 3, 4, 6) to the same set of labour market states which could play a role of $\pi_{00}^{(h)}$ estimate[oo]. Such rate is 0.9229 in the sample. The bias (possibly going both ways) is easily explained by the fact that health can be concealed in both the origin labour market state and the destination labour market state of the presumably healthy transition. Thus, the technique could only be used to substantiate the last two rows of the $\pi^{(h)}$ matrix while the estimation of the first row is only possible with the structural model.

The transition probability matrix for matching reveals the following notable patterns. About 67% of the unemployed ($m_t = 0$) are likely to remain with no full-time job offer next year. It is more probable to get a non-AFP job other than the AFP job. Both types provide reasonable job security, but on average people leave AFP jobs much more rarely (by about 4.5% points). If they do, switching to the opposite job type is 5 times more probable than becoming unemployed. Again, employees at the AFP companies are holding on to their jobs to a considerably higher degree. If this pattern holds in place for all ages of all generations, the limiting distribution assigns 37.15% and 59.08% of the population correspondingly to the non-AFP and AFP employment and implies an unemployment rate of 3.77%.

The estimate of θ^{bhm} reflecting no reduction in labour market opportunities for the people receiving disability is probably an artifact of the model construction. Since it is assumed that both agents with very bad health ($h_t = 2$) and those with bad health and no match end up in the full time disability labour market state, it is only the variation in AFP verses non-AFP job match to the disabled which identifies the parameter. Obviously, this variation is not sufficient, and people with bad health and no job match could be regarded by the model as people with very bad health. The true value of θ^{bhm} is then incorporated into $\pi_{02}^{(h)} = 1 - \pi_{00}^{(h)} - \pi_{01}^{(h)}$ which in this case is somewhat overestimated. However, so far I accept the estimate of the labour market opportunities reduction parameter which could yet change in the fourth step and move on the next step of the estimation procedure.

6.3. Estimation results: step 3.

The third step of the estimation strategy described in section 6.1 contains the estimation of the preference parameters $\lambda, a, b(\overline{s}_t), c_0, ..., c_7, c_1^{(tf)}, .., c_3^{(tf)}$ which is achieved by maximization of the likelihood function (33) with respect to the parameter vector $\theta'' = (\lambda, a, b(\overline{s}_t), c_0, ..., c_7, c_1^{(tf)}, .., c_3^{(tf)})$ while keeping the transition probability parameters fixed at the values obtained in section 6.2.

[oo] This approach will be used in section 6.5 when calculating the goodness of fit measures.

The third estimation step appears to be the most challenging for many reasons. First of all, recalculation of the value function on each evaluation of the likelihood function is inevitable here, which drastically increases the running time of one likelihood evaluation compared to the previous step. Then, since no preliminary calibration was performed for the preference parameters, finding good starting values is not a trivial task in itself. In complicated structural models the likelihood function tends to be flat even in the static setups and to have multiple local extreme points. Some parameters like the discount coefficient and the risk aversion are traditionally unstable as minor changes in their values lead to large disturbances in the other. In these circumstances the choice of optimization routine is very important. The usual argument that less iterations before the convergence is attained is more convenient, is many times strengthened because more iterations and thus more function evaluations to attain convergence may increase the running time beyond the reasonable bounds and make the whole problem computationally intractable. Unfortunately, the structure of the likelihood function most of all complicated by the choice set dependencies is so complex that finding analytical derivatives of the likelihood function with respect to the parameters was infeasible. Therefore only optimization methods based on numerically estimated derivatives could be used in spite of their generally lower convergence speed compared to the methods with analytical derivatives. In addition, using numerical methods for likelihood function maximization lead to some inaccuracies in the calculation of Hessian and sometimes troubles in calculation of standard errors for the coefficient estimates.

Nevertheless, after numerous runs of the optimization routine with slightly different settings, the final estimates had appealing properties. To search for promising starting values for the full scale optimization, I followed the strategy of fixing most of the parameters first and then adding them one by one into the optimization routine. As a result, Table 10 and Table 11 present 5 consecutive estimation approaches which lead to the final set of estimates.

As shown in Table 10 I start with simplest setup where most of the parameters take zero values, the discount factor is fixed at 0.95 and the Box-Cox parameter is fixed at 0.5. The only two preference parameters are constants with income and leisure which represent the average individual in the sample. Although the coefficients are estimated rather sharply, the sign of the leisure coefficient is ambiguous. I interpret the negative utility of leisure coefficient estimate as a result of heterogeneity which can be accounted for by introducing the additional labour market specific utility component (similar to (Rust and Phelan, 1997)). Doing so solves the problem with the negative coefficient of leisure but the income coefficient is drastically reduced. Since the income and leisure scale coefficients are reflecting the relative importance of either of these two major utility determinants, I turn back to the leisure coefficient trying to counteract the unexpected change with counting for more heterogeneity. The third set of estimates reveals the systematic difference in attitude towards leisure among healthy and not healthy individuals thus clarifying that the lack of this dimension in the utility function in the previous model caused the income coefficient to be estimated so low[pp].

Table 10. Estimates of the preference parameters at step 3. First approaches.

Parameter description			Income and leisure constant parameters only		Introduce additional utility of labour market states		Introduce heterogeneity in leisure preferences	
			Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Discount factor		β	0.95	fixed	0.95	fixed	0.95	fixed
Utility of income	Risk aversion parameter	λ	0.50	fixed	0.50	fixed	0.50	fixed
	Income scale parameter	a	0.30425	0.00254	0.06984	0.00313	0.15111	0.00376
Utility of leisure	Constant	$b(\overline{s}_t)$	-1.73200	0.00580	2.22813	0.00000*	4.36834	0.00000*
	$h_t = 1$						18.14427	0.26597
	sp_t						0.05621	0.01935
	gender						-0.66735	0.01698
Labour market state specific add. utility	OLM	c_0			9.16281	0.00000*	9.17396	0.00000*
	Pension	c_1			9.70749	0.00000*	9.66068	0.00000*
	Disability (DI)	c_2			6.65482	0.00000*	8.80819	0.00000*
	Unemployment	c_3			9.51139	0.00000*	10.07589	0.00000*
	Non-AFP employment	c_4			9.96042	0.00000*	10.42392	0.00000*
	Non-AFP employment with partial DI	c_5			14.22174	0.00000*	17.45291	0.00000*
	AFP employment	c_6			9.99934	0.00000*	10.43293	0.00000*
	AFP employment with partial DI	c_7			14.15713	0.00000*	16.79323	0.00000*
Termination function		$c_i^{(tf)}$	0.00	fixed	0.00	fixed	0.00	fixed

The last estimates also suggest higher marginal utility of leisure for the individuals from full households and lower marginal utility of leisure for women (which may reflect their tendency to work more compared to men in order to counteract, for example, the gender gap in pension benefits). Again all the coefficients have very small standard errors, although one problem is emerging. Zero standard errors (market with *) of the labour market specific coefficients are due to the inaccuracy in the numerical approximation of the Hessian[qq]. This implies that true errors of

[pp] Provided here definition of the parameter subvector $b(\overline{s}_t)$ is a result of direct search of the parameters that could most comprehensively reflect individual heterogeneity but still keep the problem computationally tractable.

[qq] Information equality is used to calculate asymptotic standard errors of the coefficients from the Hessian evaluated in the maximum point, see Greene (2000) for more details.

these coefficients fail to be calculated (although in other preliminary and not reported optimization runs they were assessed and appeared to be as small as the errors of other coefficients). Another problem with the estimations presented in Table 10 is the failure to estimate the coefficients of the termination function. In some runs and for some starting points I could obtain estimates for the termination function parameters, but in whole these parameters appear to be extremely unstable and sensitive to the changes in the income coefficient.

Table 11. Estimates of the preference parameters at step 3. Final approaches

Parameter description			Introduce discount factor and Box-Cox parameter			Combine similar labour market state dummies (the final approach)		
			Estimate	Std. Err.	p-value	Estimate	Std. Err.	p-value
Discount factor		β	0.80101	0.00088	0.00000	0.91305	0.00077	0.00000
Utility of income	Risk aversion parameter	λ	0.71885	0.01901	0.00000	0.66743	0.03031	1.7E-107
	Income scale parameter	a	0.34830	0.00505	0.00000	0.16606	0.00420	0.00000
Utility of leisure	Constant	$b(\overline{s}_t)$	0.00		fixed	0.00		fixed
	$h_t = 1$		19.46889	0.37787	0.00000	17.43456	0.15031	0.00000
	sp_t		0.38676	0.02437	9.70E-57	0.28167	0.02059	1.28E-42
	gender		-1.14443	0.02202	0.00000	-0.54832	0.01813	7.7E-201
Labour market state specific add. utility	OLM	c_0	9.03299	0.00000*	0.00000	10.54470	0.00000*	0.00000
	Pension	c_1	9.73344	0.00000*	0.00000	11.21128	0.00000*	0.00000
	Disability (DI)	c_2	8.41292	0.00000*	0.00000	5.10863	0.00000*	0.00000
	Unemployment	c_3	9.77845	0.00000*	0.00000	11.59163	0.00000*	0.00000
	Non-AFP employment	c_4	9.61362	0.00000*	0.00000	11.07245	0.00000*	0.00000
	Non-AFP employment with partial DI	c_5	18.56178	0.00000*	0.00000			$= c_7$
	AFP employment	c_6	9.71429	0.00000*	0.00000			$= c_4$
	AFP employment with partial DI	c_7	18.23398	0.00000*	0.00000	15.62574	0.00000*	0.00000
Termination function		$c_i^{(tf)}$	0.00		fixed	0.00		fixed

Unfortunately, the two described problems continue into Table 11 which displays the two remaining approaches including the final estimation result of the step 3. As anticipated the parameters of the termination function appeared to be even more sensible to the changes in the risk aversion parameter λ so that each time the optimization was performed with respect to these parameters simultaneously, the routine reported an inability to make a feasible step in the direction of the function increase. Apparently, the reason for such behavior of the likelihood function lies in the insufficient variability of the available state variables at the termination age, and is likely to be a direct consecutive of the leaving the savings process out of the model.

Otherwise the estimation results presented in Table 11 are very appealing. The two setups differ in the combined additional utilities for some of the labour market states, and even though this measure does not resolve the problem of noisy Hessian approximation, both approaches allow for sensible estimation of the discount factor and risk aversion parameter[rr]. The final estimation approach suggests the discount factor 0.91305 or the intertemporal discount rate of 9.523%; risk aversion parameter $\lambda = 0.66743$ or the risk aversion coefficient of 0.3227. The former estimate seems to be in line with the previous empirical findings whereas the latter is unusually small indicating that the observed individuals are rather risk neutral with respect to their disposable household income. Estimated in the final approach coefficients as before appoint higher marginal utility of leisure to the unhealthy men and with spouses.

Comparison of the additional labour market state specific utility components (which demonstrate rather stable behavior throughout the approaches) reveals the underlying utility patterns. The least attracting labour market state is full time disability, especially when it results from a permanent health shock (when $h_t = 2$ and no additional utility is attached to leisure). The most attractive is combined disability and employment – this reflect large alternative cost of working for those who choose to become employed while on disability. Out of the labour market state is the least attractive among the left states, while employment, unemployment and retirement result in approximately the same additional utility for the individuals occupying these states. However, if these three are compared, the unemployment state seems to be a bit more attractive than pension, and pension a bit more attractive than employment which can be interpreted as evidence for the preferences outside the income-leisure dimensions that effect the choices of the labour market states.

6.4. Estimation results: step 4.

As mentioned in section 6.1 step 4 of the estimation strategy is strictly speaking unnecessary. The estimation results contained in the last three columns of Table 9 and Table 11 already bear the desirable asymptotic properties of consistency and asymptotic normality although lacking efficiency. (Rust, 1994) provides formulas for calculation of the standard errors of the estimates from the partial likelihood Hessian matrixes. On the other hand, (Rust, 1994) suggests a simple procedure to obtain the efficient estimators which essentially is a single step of the Newton optimization algorithm. I take this approach a little bit further and perform three Newton steps from the best estimation found so far using the numerically approximated Hessian.

[rr] It could be noted that most of the times these parameters are fixed in the dynamic programming modeling papers, see (Berkovec and Stern, 1991; Gilleskie, 1998; Karlstrom et al., 2004; Burkhauser et al., 2004; Heyma, 2004).

Table 12. Final estimates of all the parameters (step 4).

Parameter description			Previous estimate		Final estimate		
			Estimate	Std. Err.	Estimate	Std. Err.	p-value
Health transition probabilities	Good to good	$\pi_{00}^{(h)}$	0.97073	0.00011	0.97087	0.00011	0.00000
	Good to bad	$\pi_{01}^{(h)}$	0.02806	0.00011	0.02779	0.00011	0.00000
Matching transition probabilities	No to non-AFP	$\pi_{01}^{(m)}$	0.19737	0.00151	0.18709	0.00148	0.00000
	No to AFP	$\pi_{02}^{(m)}$	0.13229	0.00131	0.12542	0.00127	0.00000
	N-AFP to n-AFP	$\pi_{11}^{(m)}$	0.89112	0.00045	0.89029	0.00046	0.00000
	Non-AFP to AFP	$\pi_{12}^{(m)}$	0.08915	0.00042	0.08908	0.00042	0.00000
	AFP to non-AFP	$\pi_{21}^{(m)}$	0.05584	0.00026	0.05585	0.00026	0.00000
	AFP to AFP	$\pi_{22}^{(m)}$	0.93550	0.00027	0.93502	0.00028	0.00000
Matching limitation		θ^{bhm}	0.99996	0.00001	0.99999	0.00001	0.00000
Motion rule for e_t	Constant term (1)	$c_1^{(e)}$	-0.31837	0.05124	-0.31571	0.05129	7.50E-10
	$nw_t = 9$	$c_2^{(e)}$	1.24310	0.10514	1.24059	0.10513	0.00000
	$nw_t = 10$	$c_3^{(e)}$	2.36392	0.05245	2.36119	0.05250	0.00000
	Constant term (2)	$c_4^{(e)}$	-0.30423	0.01949	-0.30211	0.01949	0.00000
	Gender	$c_5^{(e)}$	-0.62980	0.02928	-0.63135	0.02929	0.00000
Motion rule for nw_t	$nw_{t-1} = 0$	$c_1^{(nw)}$	-5.78499	0.02122	-5.78533	0.02121	0.00000
	$ps_t \in \{3,4,6\}$	$c_2^{(nw)}$	5.64890	0.01415	5.64858	0.01415	0.00000
	$ps_t \in \{5,7\}$	$c_3^{(nw)}$	13.73441	1.44000	15.66852	3.78400	3.46E-05
	$ps_t \in \{0,2\}$	$c_4^{(nw)}$	-5.93205	0.09033	-5.93309	0.09032	0.00000
	Constant term	$c_5^{(nw)}$	-0.47464	0.00863	-0.47346	0.00864	0.00000
Discount factor		β	0.91305	0.00077	0.91235	0.00091	0.00000
Utility of income	1 - risk aversion	λ	0.66743	0.03031	0.67393	0.03001	0.00000
	Constant	a	0.16606	0.00420	0.17147	0.00426	0.00000
Utility of leisure	$h_t = 1$	$b(\overline{s}_t)$	17.43456	0.15031	29.32242	0.57728	0.00000
	sp_t		0.28167	0.02059	0.26551	0.02073	0.00000
	gender		-0.54832	0.01813	-0.61363	0.01838	0.00000
Labour market state specific additional utility	OLM	c_0	10.54470	0.00000*	10.59219	0.00000*	0.00000
	Pension	c_1	11.21128	0.00000*	11.28129	0.00000*	0.00000
	Disability (DI)	c_2	5.10863	0.00000*	3.81063	0.00000*	0.00000
	Unemployment	c_3	11.59163	0.00000*	11.32202	0.00000*	0.00000
	Employment	c_4	11.07245	0.00000*	11.13257	0.00000*	0.00000
	Employment combined with disability	c_7	15.62574	0.00000*	18.24670	0.00000*	0.00000
Termination function		$c_i^{(tf)}$	0.00	fixed	0.00	fixed	

The results of the step 4 estimation are given in the last three columns of Table 12. The presented parameter estimates can be thought of as a result of applying standard maximum likelihood method in which the likelihood maximization was started from a point very close to the maximum point. Thus, the presented estimates are consistent, asymptotically normal and efficient. As it follows from the last column of the table, all the estimates are significantly different from zeros and are in general very accurately estimated.

Comparison of the final estimates to the previous ones (middle columns in Table 12) reveals little change. The most movement is associated with the marginal utility of leisure for the unhealthy individuals. Considerable increase of the coefficient is counteracted by the corresponding labour market state specific parameters. Such behavior of these estimates indicates relative flatness of the likelihood with respect to these parameters and corresponding difficulty of estimation. However, the elusive influence of the health process seems to be captured and well identified which follows from the low standard errors of the estimates.

The final estimates imply transition probabilities for health and match process very similar to (51). Slight corrections correspond to the increase in the good health to the worst health transition probability (by 0.00013) and the increase of the no match to no match transition probability (by 0.01715). The latter is related to slightly higher natural unemployment of 4.163% (vs. 3.77% previously).

The labour market state specific additional utilities are practically identical to those given in section 6.3, except that the coefficients of full time and part time disability change in reaction to the mentioned shift in the marginal utility of leisure for the potentially disabled. The values corresponding to the employment, unemployment and retirement states compress, but the structure of the additional utilities is preserved. Unfortunately, the standard errors of the labour market specific coefficients still could not be identified due to the conditions discussed in section 6.3.

The final estimate for the discount factor is 0.91235, which implies 9.607% discounting of the future expected utilities by the almost risk neutral agents (risk aversion coefficient is estimated on the level 0.67393).

Overall, the transition probabilities and preference parameter estimates obtained in the last step of the estimation and presented in the last three columns of Table 12 (apart from the skipped termination function parameters) appear to be informative, reasonable and accurate. To assess the performance of the model as whole in the next section I turn to calculating its goodness of fit.

6.5. Goodness of fit

In the literature there is a considerable diversity in the measures of goodness of fit of the structural dynamic programming models. First of all, this is a consequence of a general theoretical difficulty of applying principles of measuring goodness of fit by the means of summed squared errors (as in the linear regression analysis) for models with discrete dependent variables. The mainstream approach here (as described by (Maddala, 1988) and (Greene, 2000)) is constructing fit measures based on the likelihood ratio test statistic which implements a comparison between actual model and its "restricted" or simplified version to understand how much additional information is gained. After proper scaling these likelihood ratio statistics can be interpreted as pseudo-R^2 resembling a percentage of data variability explained by the model.

A popular pseudo- R^2 measure suggested by McFadden and used in many papers (for example, (Jia, 2005) and (Heyma, 2004)) has the form

$$R^2_{McF} = 1 - \frac{\ln\left(L(\theta^*)\right)}{\ln\left(L(\tilde{\theta}^*)\right)}, \tag{52}$$

where $L(\theta^*)$ and $L(\tilde{\theta}^*)$ are the maximum values of the likelihood function on the sets of correspondingly unrestricted and restricted coefficient vectors θ and $\tilde{\theta}$. It is easy to see that R^2_{McF} belongs to the unitary interval and approaches unity when the model attains perfect fit (with $L(\theta^*) = 1$). When the model does not differ from the benchmark case, that is when $L(\theta^*) = L(\tilde{\theta}^*)$, McFadden's pseudo-$R^2$ becomes zero.

Therefore, the choice of the restricted version of the model is very important in measuring the fit by the means of pseudo-R^2. The two simplest approaches that are used most of the time are restricting parameter vector to constants or restricting the parameter vector to a single value, namely zero. Both approaches can be well interpreted in the static discrete choice models – the latter corresponds to uniform distribution of choice probabilities, in this case the denominator in (52) has a particularly simple form of $-n \cdot \ln(m)$, where n is number of individuals and m is number of choices. The former approach can be shown to be identical to substituting the choice probabilities in the likelihood function with the observed frequencies $\frac{n_i}{n}$, where n_i denotes a number of individuals observed to choose an alternative *i*.

In a dynamic context[ss], however, standard restrictions for the parameter vector fail to facilitate the described interpretations due to the use of value functions instead of simple utilities in the formula (26) for choice probabilities. That is why the majority of papers estimating structural dynamic models skip the unifying likelihood ratio approach and instead use distributed measures based on comparison of some predicted values with the ones observed in the data. Plotting the differences on the appropriate axes leads to mostly used graphical representation of goodness of fit found in (Karlstrom et al., 2004), (Burkhauser et al., 2004), (Gilleskie, 1998), (Heyma, 2004), (French, 1999) and section 5.2 of the current paper. Mostly compared are predicted choice probabilities and the observed frequencies of choices (like in (Rust and Phelan, 1997)) which is possible because the panel data required for estimation of a dynamic structural model most often allows for calculation of such frequencies. As described in (Maddala, 1988) this is the feature of grouped data, and a standard Pearson chi-square test can be applied to make the comparison formal (as it is done in (Gilleskie, 1998) and (Gustman et al., 2002)).

An even more detailed approach suggested by McFadden and described in (Maddala, 1988) requires tabulating the pairs of observed and simulated outcomes (choices) for each observation and constructing some aggregating measures on the resulting matrixes. This approach is not found in the dynamic literature probably because of the additional complication and artificialization of this approach by the repeated nature of the dynamic choice.

In the present paper along with the graphical representation of fit I use the concept of McFadden's likelihood ratio adopting it for the dynamic settings. As it follows from the description above, the only adaptation needed refers to the specification of the benchmark version of the likelihood function. Instead of restricting the parameter vector in the benchmark case I compute the following proxies for the choice probabilities:

$$P_t^{freq}(d_t \mid s_t) = \frac{N\left\{(d_t^a, s_t^a) : d_t^a = d_t, s_t^a = s_t\right\}}{N\left\{(d_t^a, s_t^a) : s_t^a = s_t\right\}}, \tag{53}$$

$$\forall t : P^{freq}(d \mid s) = \frac{N\left\{(d_t^a, s_t^a) : d_t^a = d, s_t^a = s\right\}}{N\left\{(d_t^a, s_t^a) : s_t^a = s\right\}}, \tag{54}$$

$$P^{uniform}(d_t \mid s_t) = \frac{1}{\left|D(s_t)\right|}, \tag{55}$$

where the notion $N\{\bullet\}$ denotes the number of elements in the collection described by the properties in the brackets and the notion $|\bullet|$ denotes the number of elements in a set. The first two

[ss] When discount factor β does not equal 0.

approximates for the choice probabilities are based on the observed frequencies of choices in each point s_t of the state space and differ in the treatment of time. $P_t^{freq}(d_t \mid s_t)$ are calculated as age-specific values whereas $P^{freq}(d \mid s)$ are averaged over all ages. The reason for calculating both of these frequency based proxies is the possibility of the lack of observations for some values of the state vector s_t in some ages. Aggregation over ages can level off this possible problem although on the obvious expense of loosing accuracy of the estimate. For those values of s_t when denominators in (53) or (54) become zeros, I use the simple uniform choice probability distribution formula (55) which assigns equal probabilities to all possible choices in the relevant choice set $D(s_t)$.

Either of the approximations of the choice probabilities (53)-(55) may be used in a standard structural dynamic model for calculating restricted likelihood function which in turn may be used in formula (52) for computing McFadden's likelihood ratio fit measure. Still, in the present model one difficulty remains. Recall that vector $s_t^a = \left(ps_t^a, h_t, m_t, e_t^a, sp_t^a, nw_t^a, aw_t^a \right)$ contains two latent variables h_t and m_t which are integrated out when calculating the likelihood function. The values of these latent variables also have to be estimated[tt]. I apply the following simple approximation rules which allow reconstruction of the health and match variables given the values of the rest of the state variables (namely, the labour market state in the current period).

$$\begin{aligned} h_t^a &= 0 \Leftarrow ps_{t+1}^a \notin \{2,5,7\}, \\ h_t^a &= 1 \Leftarrow ps_{t+1}^a \in \{5,7\}, \\ h_t^a &= 2 \Leftarrow ps_{t+1}^a = 2, \end{aligned} \tag{56}$$

$$\begin{aligned} m_t^a &= 0 \Leftarrow ps_{t+1}^a \in \{0,1,2,3\}, \\ m_t^a &= 1 \Leftarrow ps_{t+1}^a \in \{4,5\}, \\ m_t^a &= 2 \Leftarrow ps_{t+1}^a \in \{6,7\}. \end{aligned} \tag{57}$$

Application of the formulas (56) and (57) allows for the calculation of the frequency estimates for the choice probabilities in the model.

The last complication relevant for the dynamic models deals with approximating the transition probabilities and imposing proper restrictions on the corresponding parameters (these parameters and probabilities are not present in the static case at all). Following the logic of the frequency and uniform approaches, I consider three approximations for the transition probabilities: optimal values (as estimated within the maximization of likelihood for the whole model), frequency based

---

[tt] It should be possible to integrate them out of the frequency expressions as well as the likelihood function, but for the purpose of the constructing the benchmark case for the goodness of fit measures the simple approximation seems adequate enough.

approximation and uniform distribution. Intuition under these three approximations for the transition probabilities is straightforward. Using optimal values for the corresponding parameters allows for assessment of the amount of variation explained by structural modeling of choices as opposed to simple modeling of transitions. Frequencies based approximation is equivalent to re-estimating the motion rules for individual AFP eligibility indicator *e* and the number of consecutive years with sufficient wage income *nw* (see Table 5 and Table 6) with the constant parameters only; and using frequencies of health and match transitions instead of parameterized Markovian transition probabilities (as in (15) and (16))[uu]. The rest of the state variables evolve according to the motion rules fixed within the model: stochastic (like spouse indicator *sp*) or deterministic (like labour market state *ps* and aggregate wage *aw*) which do not require neither estimation nor, thus, approximation. Finally, uniform distributions for the transition probabilities can be easily implemented by setting the parameters in the motion rules for *e* and *nw* to zeros and using equal probabilities of transitions for health and job match. With respect to the latent variables this last approximation can be interpreted as skipping the health and match processes completely out of the model.

The described three approximations for the choice probabilities and the three approximations for the transition probabilities can be conveniently presented in a table (see Table 13). Here nine versions of McFadden's pseudo-R^2 are accompanied by the corresponding maximum values of restricted likelihoods calculated as the likelihood function (33) in which choice probabilities $P_t(d_t^a \mid s_t^a, \theta)$ and transition probabilities $p(s_t^a \mid s_{t-1}^a, d_{t-1}^a, \theta)$ are substituted by the corresponding approximations. The initial condition probabilities $p_0(m_{T_0^a}, h_{T_0^a}, \theta)$ are also effected by the approximated parameters of transition probabilities.

Table 13 displays quite a variety of pseudo-R^2 ranging from modest 9.12% to rather outstanding for discrete choice analysis 79.064% depending on the calculation technique. Analysis of the different obtained values reveals certain patterns. First, as the columns of the table completely dominate each other, that is the largest R_{McF}^2 in one column is still smaller than the smallest R_{McF}^2 in the other for each pair, it is obvious that transition parameters play a more important role in the overall model fit. This is also supported by the low fit measures in the first column – the amount of additionally explained variation is rather small if the benchmark case is based on the optimal transition probabilities. Second, the difference between more accurate age specific approximations of the choice probabilities and the less accurate aggregates over all ages is much more pronounced for the

[uu] Here frequencies for the latent variables are calculated on the basis of their approximated values from (56) and (57), thus leading to double approximation. Still, it seems sufficient for goodness of fit measures.

true values of transition probabilities and completely diminishes when the transition probabilities are roughly guessed. Thus, the static principle of using observed frequencies of choices as the bases for the goodness of fit measurement is quite applicable in the dynamic models in spite of some arising ambiguity. Third, comparison of the obtained values of R^2_{McF} in the third column with the corresponding values from the first column clearly shows the importance of latent variables of health and job match introduced in the model since destroying their transition mechanism changes the fit more than dramatically.

Table 13. Goodness of fit: McFadden's likelihood ratio indexes.

Choice probability restriction	Transition probability restriction		
	No restriction, optimal values	Frequency based approximation	Uniform approximation
Frequency based age specific approximation $P_t^{freq}(d_t \mid s_t)$	$R^2_{McF} = 9.120\%$ $L(\tilde{\theta}^*) = -1460907.66617$	$R^2_{McF} = 47.895\%$ $L(\tilde{\theta}^*) = -2548076.47686$	$R^2_{McF} = 78.030\%$ $L(\tilde{\theta}^*) = -6043159.71612$
Frequency based aggregated over age approximation $P^{freq}(d \mid s)$	$R^2_{McF} = 18.939\%$ $L(\tilde{\theta}^*) = -1637875.96375$	$R^2_{McF} = 53.129\%$ $L(\tilde{\theta}^*) = -2832605.20954$	$R^2_{McF} = 77.729\%$ $L(\tilde{\theta}^*) = -5961513.80777$
Uniform approximation $P^{uniform}(d_t \mid s_t)$	$R^2_{McF} = 41.823\%$ $L(\tilde{\theta}^*) = -2282118.52527$	$R^2_{McF} = 61.990\%$ $L(\tilde{\theta}^*) = -3492976.15806$	$R^2_{McF} = 79.064\%$ $L(\tilde{\theta}^*) = -6341716.39357$
Optimal unrestricted likelihood value	$L(\theta^*) = -1327676.68661$		

To point out one measure of goodness of fit from the Table 13 I again appeal to the static case where frequency based approximation and uniform approximations are applied most often. Keeping in mind the Bayesian tradition it seems appropriate to use the former approach as a way to utilize available information whereas the latter approach suits the situation when no prior information is available. Thus, since it is generally possible to compute frequencies for the choice probabilities and choose the easier and more robust way to compute them, I appeal to the age aggregate frequency based approximation $P^{freq}(d \mid s)$ for the choice probabilities; whereas approximation of transition probabilities may be approached as if there was no prior information (third column) or through frequencies of approximated health and match as demonstrated (second column). Since the job match is recoverable from the labour market state data, the latter approach seems more grounded in the present model. Thus, the final measure of goodness of fit should be 53.129%.

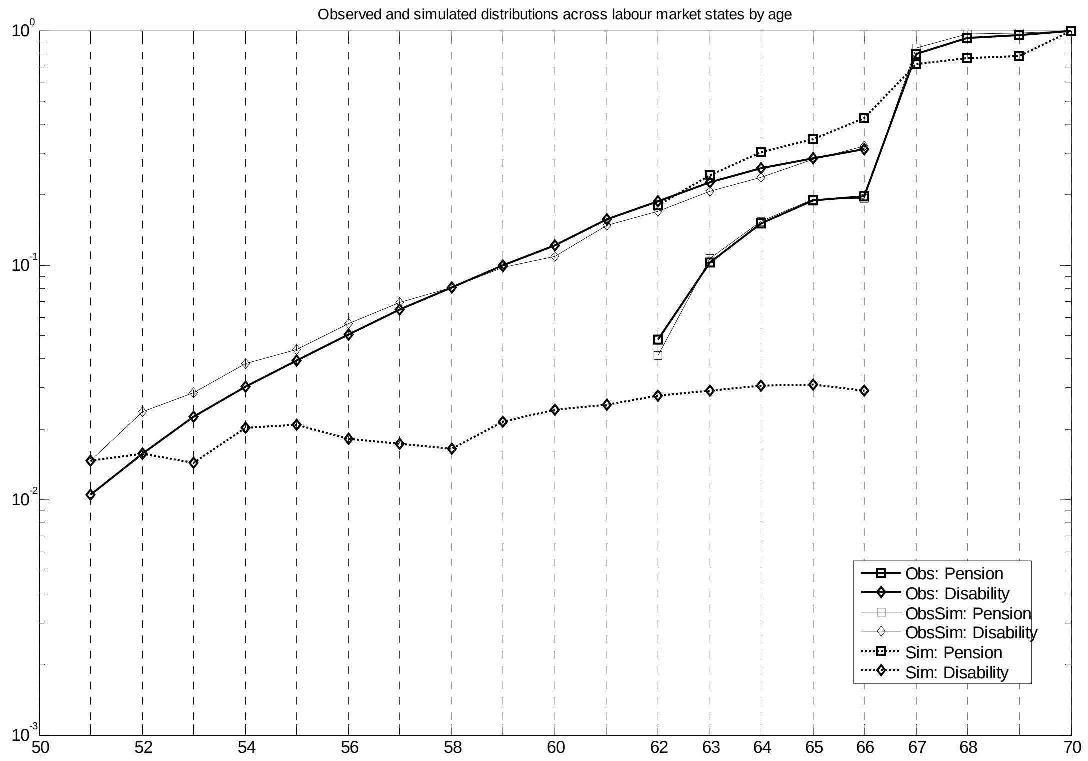

Figure 5. Observed and predicted fractions of pension and disability labour market states.

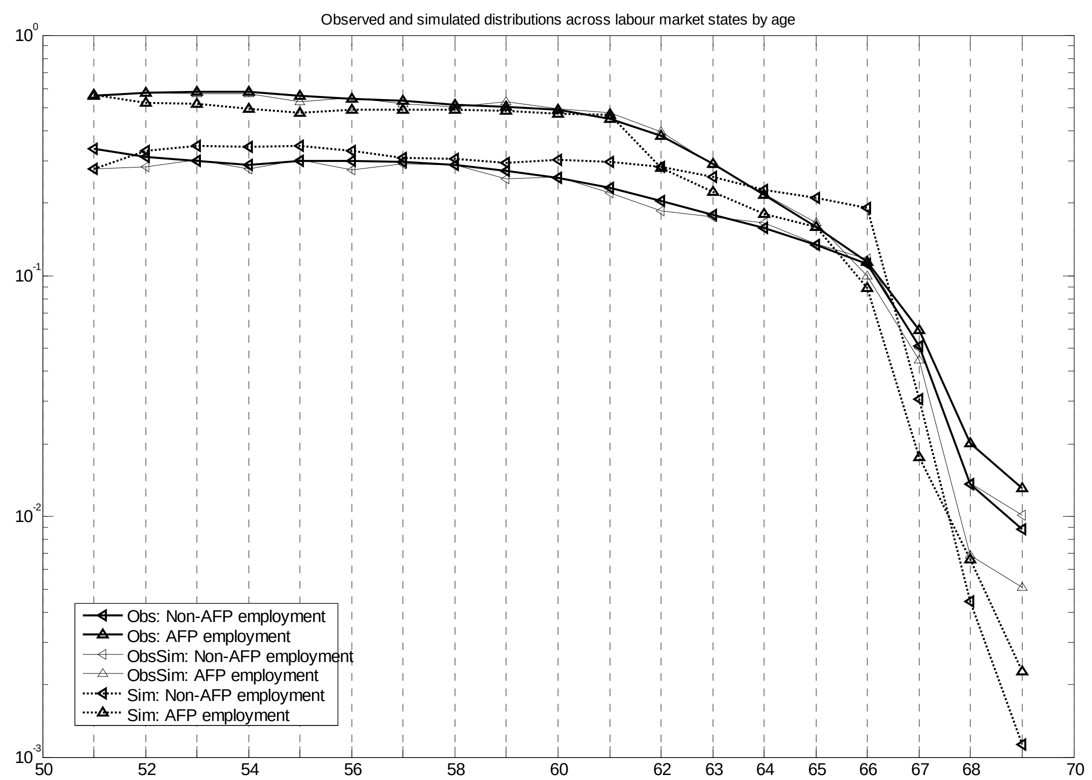

Figure 6. Observed and predicted fractions of employment labour market states.

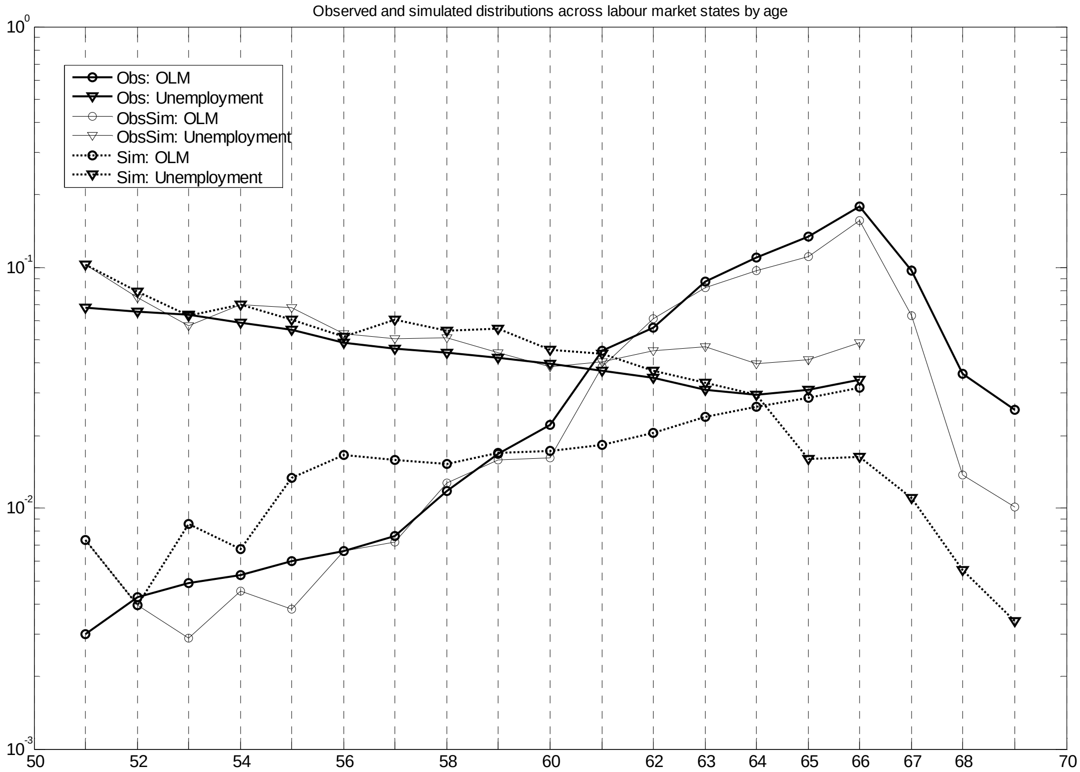

Figure 7. Observed and predicted fractions of labour market states.

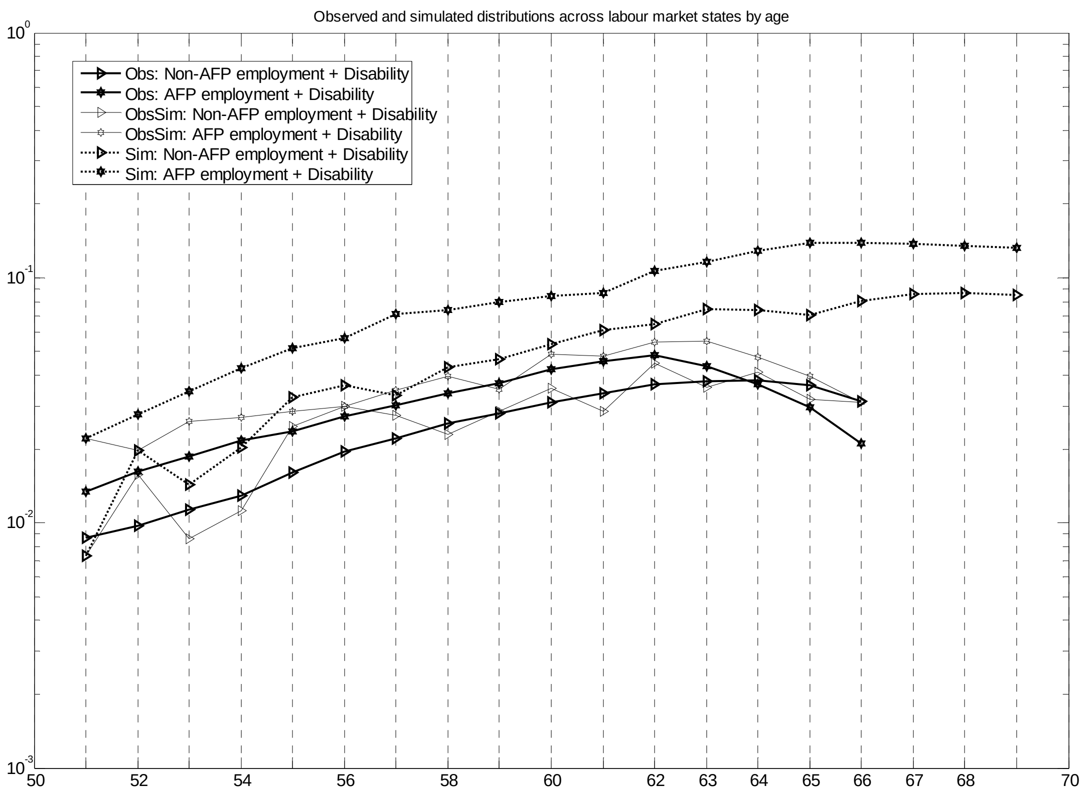

Figure 8. Observed and predicted fractions of partial disability and employment states.

Table 14. Observed distributions of states on the labour market by age.

State	50	51	52	53	54	55	56	57	58	59
0	0.0000	0.0025	0.0046	0.0056	0.0061	0.0069	0.0076	0.0090	0.0129	0.0182
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0088	0.0172	0.0290	0.0429	0.0583	0.0755	0.0967	0.1203	0.1464
3	0.0665	0.0626	0.0601	0.0579	0.0540	0.0502	0.0453	0.0420	0.0390	0.0369
4	0.3649	0.3495	0.3281	0.3127	0.2942	0.2992	0.2965	0.2891	0.2795	0.2604
5	0.0129	0.0128	0.0147	0.0179	0.0207	0.0252	0.0293	0.0327	0.0362	0.0389
6	0.5412	0.5483	0.5583	0.5575	0.5597	0.5360	0.5186	0.5007	0.4787	0.4630
7	0.0146	0.0154	0.0170	0.0196	0.0223	0.0242	0.0272	0.0298	0.0333	0.0362
60	61	62	63	64	65	66	67	68	69	70
0.0233	0.0436	0.0528	0.0808	0.0991	0.1191	0.1580	0.0832	0.0288	0.0198	0.0000
0.0000	0.0000	0.0454	0.0962	0.1382	0.1700	0.1758	0.8208	0.9417	0.9612	1.0000
0.1743	0.2071	0.2417	0.2828	0.3199	0.3520	0.3804	0.0000	0.0000	0.0000	0.0000
0.0335	0.0314	0.0289	0.0256	0.0242	0.0251	0.0277	0.0000	0.0000	0.0000	0.0000
0.2390	0.2167	0.1910	0.1655	0.1460	0.1238	0.1029	0.0451	0.0119	0.0074	0.0000
0.0420	0.0440	0.0466	0.0470	0.0460	0.0422	0.0356	0.0000	0.0000	0.0000	0.0000
0.4470	0.4130	0.3470	0.2603	0.1911	0.1395	0.0992	0.0509	0.0176	0.0116	0.0000
0.0408	0.0443	0.0465	0.0418	0.0354	0.0283	0.0205	0.0000	0.0000	0.0000	0.0000

Table 15. Simulated distributions of states on the labour market by age (final estimation).

State	50	51	52	53	54	55	56	57	58	59
0	0.0000	0.0074	0.0040	0.0086	0.0068	0.0133	0.0166	0.0159	0.0153	0.0170
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0147	0.0158	0.0144	0.0203	0.0210	0.0183	0.0174	0.0165	0.0215
3	0.0000	0.1029	0.0791	0.0632	0.0700	0.0610	0.0515	0.0608	0.0547	0.0556
4	0.0000	0.2794	0.3320	0.3477	0.3431	0.3467	0.3306	0.3097	0.3053	0.2948
5	0.0000	0.0074	0.0198	0.0144	0.0203	0.0324	0.0365	0.0333	0.0433	0.0465
6	0.0000	0.5662	0.5217	0.5172	0.4966	0.4743	0.4900	0.4920	0.4911	0.4853
7	0.0000	0.0221	0.0277	0.0345	0.0429	0.0514	0.0565	0.0709	0.0738	0.0794
60	61	62	63	64	65	66	67	68	69	70
0.0172	0.0183	0.0206	0.0239	0.0263	0.0288	0.0315	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.1796	0.2412	0.3024	0.3451	0.4217	0.7176	0.7614	0.7760	1.0000
0.0243	0.0255	0.0279	0.0291	0.0306	0.0310	0.0293	0.0000	0.0000	0.0000	0.0000
0.0456	0.0438	0.0372	0.0333	0.0295	0.0160	0.0163	0.0110	0.0055	0.0034	0.0000
0.3040	0.2966	0.2828	0.2588	0.2287	0.2105	0.1924	0.0308	0.0044	0.0011	0.0000
0.0537	0.0612	0.0650	0.0748	0.0738	0.0705	0.0804	0.0857	0.0866	0.0848	0.0000
0.4711	0.4679	0.2807	0.2225	0.1802	0.1592	0.0891	0.0176	0.0067	0.0023	0.0000
0.0841	0.0866	0.1063	0.1164	0.1286	0.1389	0.1391	0.1374	0.1354	0.1324	0.0000

Along with McFadden's likelihood ratio based pseudo-R^2, I also draw the distributions among the labour market states for the 1000 simulated households by age – similar to the graphs in section 5.2 Table 14 and Table 15 contain the plotted data – respectively observed and simulated frequencies. As before, the initial characteristics for these hypothetical households are chosen randomly from the real data initial conditions, and the graphs represent the outcomes of the solution of the agents sequential decision problem (2). In addition to the full sample observed distributions of labour market states, I plot these distributions for the particular random sub-sample used in the initial conditions randomization.

As seen from the graphs (Figure 5-Figure 13) and the tables (Table 14 and Table 15) the model correctly replicates all the tendencies and in general presents reasonable fit. The most problematic deviations are found in prediction of disability which is reasonable in the beginning and is gradually more and more underestimated (however, differences on the logarithmic scale are a bit overemphasized). This is a possible consequence of the effects of two factors. First, the model may be too strict in the assumption of markovian evaluation of health so that introduction of age dependence in health could be fruitful. This seems plausible but not very likely because the model displayed reasonable flexibility under calibration in section 5 (compare, for example, Figure 2 and Figure 5). The second explanation seems much more realistic – the chosen estimation method relies heavily on the distributional assumptions which could be violated in the given data sample. Besides, there is also a numerical issue here: since bad health necessary for disability retirement is relatively much less probable, individual likelihood associated with disabled individuals may be "lost" in the computation of the likelihood function due to the round off problem. The latter effect is additionally worsened by the uneven data panel available for the estimation (presented graphically in Figure 12).

Correspondingly to underestimated full time disability, partial disability is systematically slightly overestimated which is a sign of measurement errors in the characteristics of the combined disability and employment labour market state. Also, AFP pension is somewhat overestimated which corresponds to a more pronounced kink in the employment curve for the AFP companies.

In general, however, the model displays reasonable fit adequately representing observed dynamics on the labour market. This allows me to move on and present one application of the estimated model.

7. Policy simulation

Completely estimated in section 6.4 and tested for the goodness of fit in section 6.5 the dynamic model of health and retirement may now be used for simulation runs in order to quantify the effects of any given changes of the environment the agents operate in. This last section present an example of such policy analysis.

7.1. Substitution between AFP and disability

AFP and disability insurance have been the two major exit routes from the labour marker in Norway for people before the normal retirement age of 67. Therefore for the policy maker the question of substitution of the two schemes becomes equivalent to the question of whether the policy aimed at the increase of labour force participation of the older workers should be designed so that it affect these schemes jointly or the issues may be regulated separately. As noted in the introduction, the question of substitution between AFP and disability have been already given attention in Norway, but the conclusions were rather different[vv].

The estimated dynamic structural model is more than suitable to answer the question of substitution. Applying a structural model for policy questions is very natural in itself, but the dynamic model is also capable of tracking not only immediate consequences of a policy, but in addition a distant (in time) reactions of the labour market and the individuals. Consider the question of potential substitution between the AFP and disability labour market state. One way to address this issue is to eliminate AFP completely and analyze the changes in the distribution of the agents across other states on the labour market. Unlike static models the dynamic model picks up the changes in behavior not only in the years affected by the simulated policy (in this case age 62-66) but also in other periods of life reflecting peoples' reevaluation of future opportunities and corresponding correction in their decision making.

Such "AFP elimination" policy is implemented into the model by simply shifting the AFP age up to 67 for all individuals. Then the model is used to simulate the responses of 1000 households in the new conditions but using the same estimates for the structural parameters. Similarly sections 5.1 and 6.5 1000 randomly chosen households from the sample provide the initial conditions for simulation, the value function is computed in the new conditions and the realizations of the decision-state process are constructed for each household to result is a set of simulated data on all the state variables including the occupied labour market states. The distribution of the latter are

---

[vv] Røed and Haugen (2003) find no substitution while Bratberg, Holmas and Thogersen (2004) find 8.6%-22.4% substitution effect (see section 1).

compared to the pre-policy simulated distribution to highlight the changes induced by the policy. Thus, the core idea under the simulation is using the same structural parameters for preferences and transition probabilities in the new circumstances implied by the policy. To make the two sets of simulated distributions of individuals among the labour market states accurately comparable, the sequence of random numbers that is used for modeling the stochastic events in the simulations is fixed between the simulation runs. It is also important that the implementation of the policy into the model happens entirely through the exogenous and not through the state variables[ww].

7.2. Simulation results

Table 16 and Figure 9 present the result in the form of the differences in the simulated distribution across the labour market stages by age (in percentage points). In the second row of the table the bold numbers show the reduction in the AFP retirement frequencies due to the introduced policy. Compared to the corresponding cells of Table 15 one can conclude that AFP pensions are eliminated completely as the policy suggests. It should be noted that since the simulated policy was aimed only at the ages from 62 to 64, the negative effects on the retirement after that age should be considered among the outcomes of the policy change. Table 16 only shows ages from 61 t0 69 due to the fact that the only response of the model is seen on this interval (in other words, all the omitted from Table 16 numbers are zeros).

It is immediately seen that the people which would otherwise be retired with the AFP pensions mostly distribute themselves across working states. Substitution into the full time disability is practically neglectable with the increase of only about one tenth of a percent point at ages 64 to 66. In the same time, the fractions of disabled employed increased much more in the same period. The bottom parts of the Figure 9 gives a clear representation of the magnitudes of the impact for different labour market states. Here only positive responses in the frequencies to the only decline induced by the policy are shown distinctly. The biggest inflow of the displaced AFP pensioners is seen in AFP employment. This is natural because exactly these employees loose the option to retire, this is also why AFP employment combined with disability is the third biggest increase. The second biggest gain corresponds to the non-AFP employment which indicates together with the unemployment that most of the displaced AFP retirees stay on the labour market. Inflow into full time disability is neglectable, but disability combined with non-AFP employment is in the third and fourth positions respectively.

[ww] In the current policy simulation the individual afp age is actually a state variable, but it only effects the choices through calculation of pension benefit and spouse income (see Table 4), the individual specific AFP age is saved in a separate variable and is used in these calculations.

Table 16. Policy simulation: elimination of the AFP scheme.

States	61	62	63	64	65	66	67	68	69
OLM	0	0	0.21	0.21	0.32	0.43	0	0	0
Pension	0	**-17.96**	**-24.12**	**-30.24**	**-34.51**	**-42.17**	-6.81	-4.44	-4.30
Disability	0	0	0	0.11	0.11	0.11	0	0	0
Unemployment	0	0.21	0.42	0.21	0.53	0.76	0.77	0.22	0.11
Non-AFP work	0	1.34	2.6	4.43	6.09	7.07	0.66	0	0
Non-AFP work + DI	0	0	0	0.11	0.32	0.65	0.88	0.89	0.68
AFP work	0	16.41	20.37	23.92	25.21	30.54	1.43	0	0
AFP work + DI	0	0	0.52	1.26	1.92	2.61	3.08	3.33	3.51

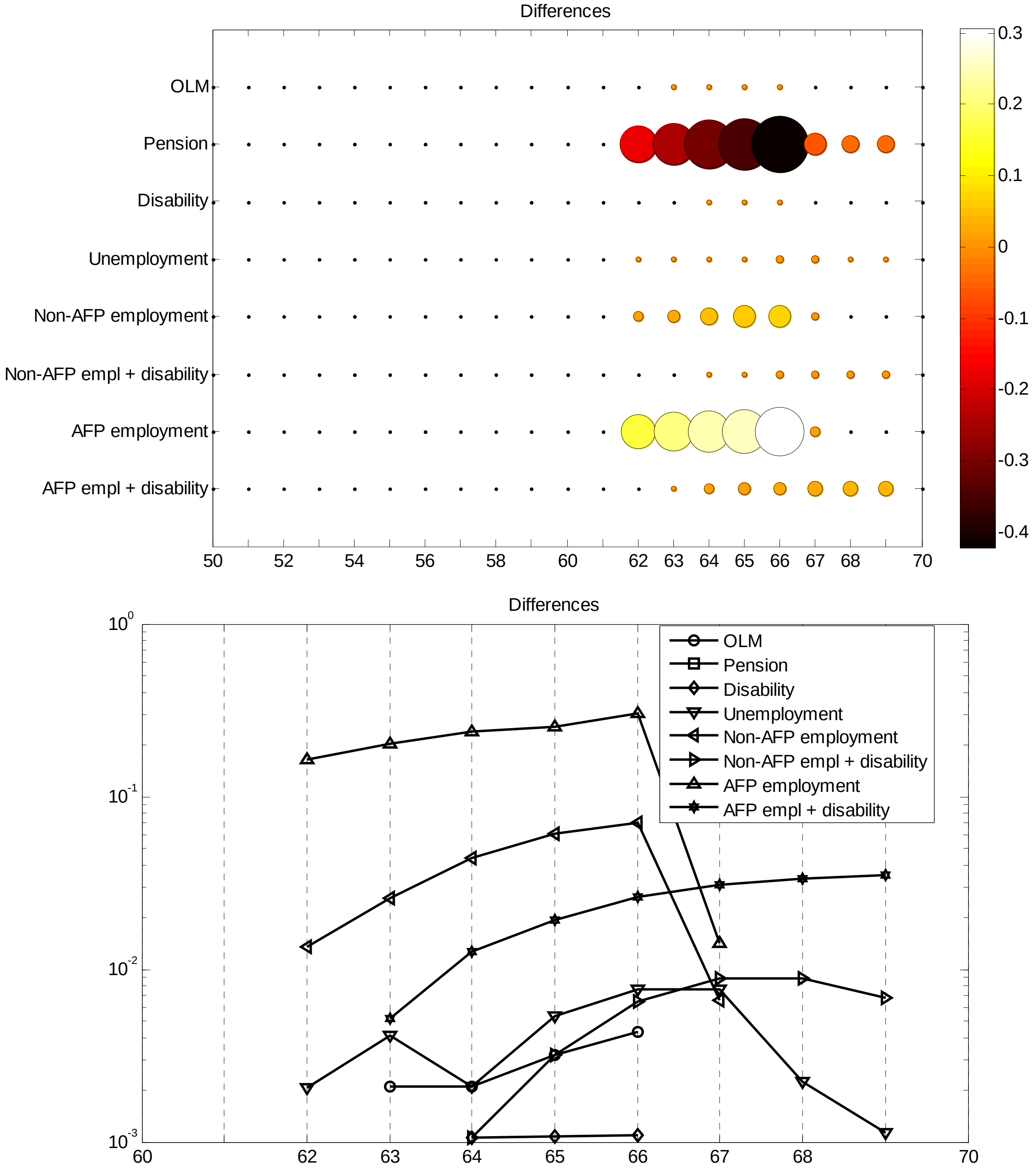

Figure 9. Graphical representation of policy simulation results.

Table 17. Policy simulation: elimination of the AFP scheme with joint disability and employment.

States	61	62	63	64	65	66	67	68	69
OLM	0	0	0.21	0.21	0.32	0.43	0	0	0
Pension	0	**-17.96**	**-24.12**	**-30.24**	**-34.51**	**-42.17**	-6.81	-4.44	-4.30
Disability	0	0 (0%)	0.51 (2.16%)	1.48 (4.88%)	2.35 (6.81%)	3.37 (7.99%)	3.96 (58.1%)	4.22 (95%)	4.19 (97.3%)
Unemployment	0	0.21	0.42	0.21	0.53	0.76	0.77	0.22	0.11
Employment	0	17.75 (98.9%)	22.97 (95.3%)	28.35 (93.7%)	31.30 (90.7%)	37.61 (89.1%)	2.09 (30.7%)	0 (0%)	0 (0%)

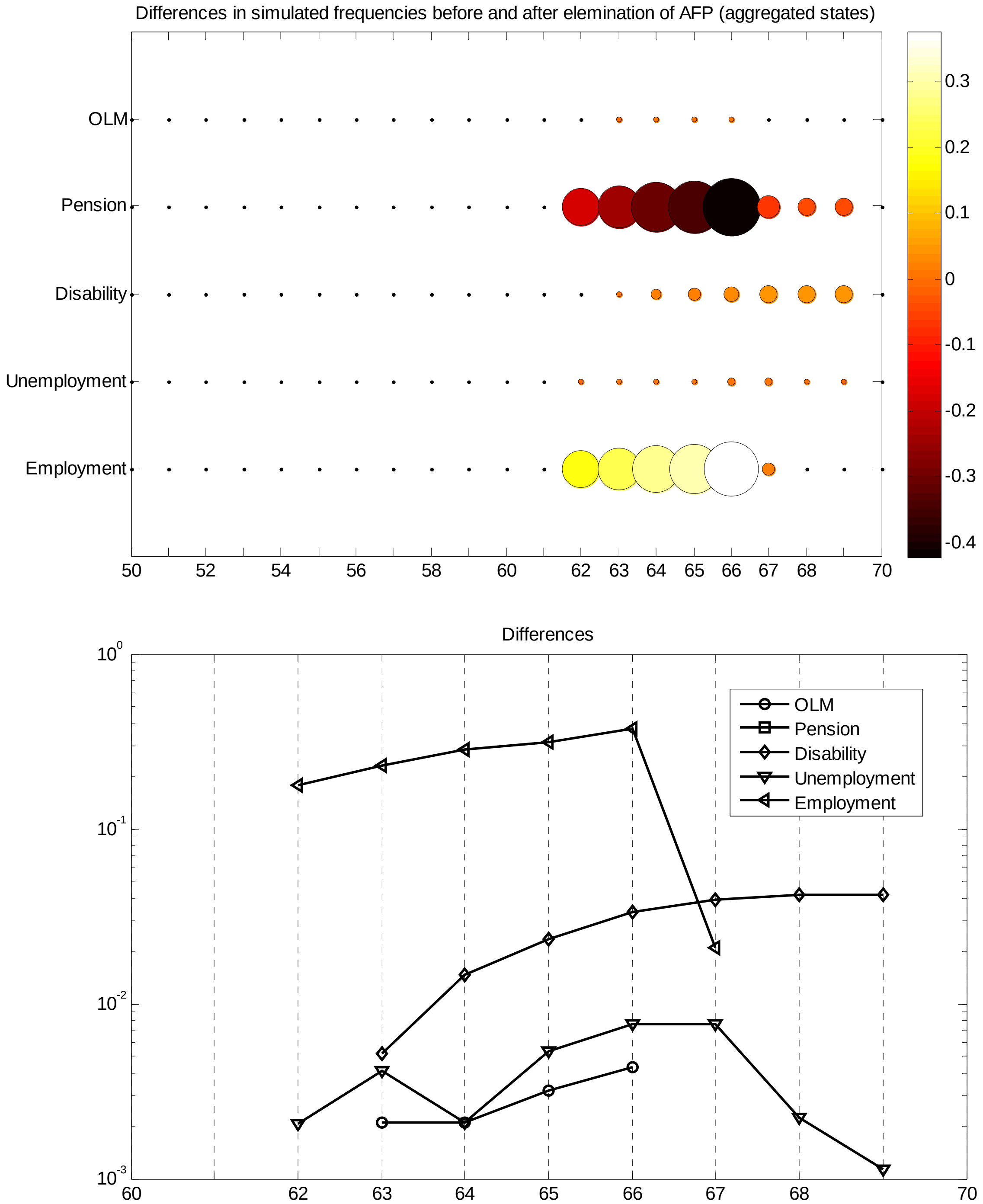

Figure 10. Graphical representation of policy simulation results (joint disability).

To make the results comparable to the previous studies, it is necessary to combine the three disability states as shown in Figure 10 and Table 17 (the two employment states are also combined). Now when disability is given a slightly different definition, the conclusions seem to change as well. From the bottom of Figure 10 it is now seen that even though the largest substitution is directed toward employment, the second largest escape for the displaced early pensioners is the disability state. These results can be interpreted as displaying signs of AFP-disability substitution after the age of 63 and up to the age of 69. Table 17 reports in prentices the fractions of the total decrease in AFP the disability and employment states are responsible for. Thus, the model predicts the substitution effect at the level of about 4.88-7.99% in the ages 64 to 66, which is intermediate between (Bratberg et al., 2004) and (Røed and Haugen, 2003) findings.

An interesting feature of the result is the absence of any reaction for the anticipated reduction of opportunities due to the elimination of AFP program. When this policy was preliminary simulated on the calibrated values of the parameters, there were strong signs for a sensible employment supply reduction starting as early as from age of 52. Now when the preference parameters are accurately estimated, it can be clearly demonstrated that the reduction of the opportunities induced by the simulated policy is not capable of causing any pre-policy behavioral responses.

The model is however picking up a dynamic response – as follows from Table 16 and Table 17 – after the age of 66. Negative numbers corresponding to the retirement state allow to conclude that the simulated policy induces general postponing of retirement in the society. Partially, this is an understandable trail of the effect that can be rationalized by the slow motions of the social processes. But the bottom of Figure 10 also indicates that while the fraction of unemployed is increasing towards the age of 67 and fades away towards 70, the fraction of disabled is increasing always while the model permits (at 70 everyone is forced into retirement by construction) although with decreasing intensity.

Overall, when AFP pension is removed, people remain on the labour market and utilize the possibility of taking out a disability pension as well. This behavior, in fact, may be interpreted as substitution into disability.

8. Conclusions

Labour market behavior among the older workers presents a complex process with several hidden factors affecting it. The current paper makes an effort to model some of these factors as latent variables evolving dynamically as Markov processes in discrete time. First, health is modeled to represent an option of taking out disability pension in order to retire or to combine disability benefits with a part-time job. Second, the matching process is introduced to differentiate between voluntary and involuntary labour market outcomes that are observed in the data. The individual AFP eligibility rules governing individual access to the early retirement option are also modeled endogenously in the dynamic context. Finally, the demographic elements as marriage and survival of both the individual decision maker and the spouse were taken into account through exogenous parameters. On the bases of these core state variables the model developed in the paper focused on careful reconstruction in each time period of the individual choice sets with regards to the decision about the future labour market position. Forecasting of four different sources of income in each of 8 different labour market states considered in the model allowed to assess structural parameters of the indirect utility function dependent on income and leisure. In the described stochastically evolving environment the decision makers react to the situation they find themselves in at the beginning of each period and try to optimize the outcome from the point of view of the expected discounted lifetime utility maximization.

The chosen structural dynamic setup proves worthwhile in analyzing the consequences of different changes in the regulations in the field of disability and retirement because it enables the model to pick up the effects spread widely in time and provides solid grounds for the analysis of behavioral responses. After the model is structurally estimated with the incomplete information maximum likelihood method, the estimates are used to address the question of substitution between the AFP retirement scheme and disability insurance. Simulation of the eliminating the AFP system altogether reveals very little substitution of the displaced AFP retirees into full time disability but provides signs of postponed retirement and disability pension take up. The observed substitution into partial disability is responsible for 5-8% of the number of misplaced AFP pensioners. This leads to a conclusion that dynamic analysis of a labour market behavior is essential and helpful tool in policy design which allows tracking long run consequences of the changes in the operating environment of the decision makers.

In general the developed model seems promising in the labour market analysis and displays great potential of the structural dynamic model on the labour market. Thus, this rather broad paper shall serve as the background for several shorter and more focused publications which will follow.

Appendix

A.1. Institutional settings of Norwegian retirement

Institutional settings that govern the retirement process in Norway are the following. All permanent residents are provided with earnings based old age public pension with defined benefits. Usual retirement age is 67 regardless of gender but it is possible to work for three more years in order to complete the full tenure of 40 years.

Old age pension consists of three major component. First, basic pension (1G) is paid to everybody with at least 3 years of working life. The level of basic pension is corrected every year at the rate exceeding CPI. Second component is an earnings based pension which is calculated as a function of earnings history expressed in basic pension amounts G. Recalculation of annual wages into pensions points is performed with a piecewise linear convex function with truncation both above and below (so that neither too small nor too large wages are taken into account). The average of 20 best pension points is then multiplied by 0.42 (0.45 if they were earned before 1992) and by *G* to give the level of the earnings based component. The third component is formed of special supplementary terms aimed on preventing the pension to go lower certain minimum.

In addition to the old age pension provided by the national social security administration so-called occupational or employer based pensions also constitute considerable support for the retirees. Occupational pensions are managed by the insurance companies and financed by the employers who are given the opportunity to deduct the payments to the system from the taxation given that the number of regulation requirements are satisfied.

An early retirement scheme (AFP) was introduced in 1989 and gave an opportunity to retire earlier than regular age with no loss in their pension benefits. The program covers the whole public sector and part of the private sector. In order to be eligible an individual must be employed in a company covered with the program and in addition meet certain individual criteria which include:

- Having been employed in the AFP-company the last 3 years or having been covered by AFP scheme during last 5 years;
- Having earnings no less than *G* prior to the year of take up and one year before;
- Not receiving pensions or similar payments from employer without work effort in return;
- Having at least 10 years after the age of 50 with earnings no less than 1*G*;
- Having the average earnings in 10 best years since 1967 no less than 2*G*.

The age of early retirement has been gradually lowered from 66 when it was initially introduced in 1989 to 65 from 1990, 64 from 1993, 63 from 1997 and finally to 62 from March 1st, 1998.

The pension level calculations under AFP scheme are aimed to provide the same pension benefit as if person would continue working until the ordinary retirement age instead of retiring early. This implies that the pension points in the years between the AFP eligibility age and 67 should be forecasted: it is done with use of the maximum between the average of the last three earned pension points and the average of ten best points from whole working history. Once the missing pension points are forecasted, the AFP pension is calculated with regular old age pension calculation technique[xx].

The disability pension roughly corresponds to the early retirement pension also compensating for the shorter job history[yy].

The tax levels are generally lower for pensioners compared to the working people and differ for single and married individuals. AFP pensions are taxed a little differently from other types of pensions. Haugen (2000) gives details on the tax function.

[xx] For more detailed see Røgeberg (2000).

[yy] For the disability program details see Bratberg (1999), Bratberg, Nilsen and Vaage (2005).

A.2. Tables

Table 18. Households in the sample.

Individuals and families in the sample	Spouse observable in the detailed data			All
	both	female	male	
Single person of full family full	87206	53164	59792	200162
single	.	57254	49198	106452
All	87206	110418	108990	306614

Table 19. Filtering report.

Reasons for filtering	Number of observations	Fraction of deleted	Fraction of all
All observations	306 614		
Incorrect initial labour market state	73 961	69.98 %	24.12 %
Static observation (only one period)	1320	1.25 %	0.43 %
Missing essential data	140	0.13 %	0.05 %
Extreme household income	17 681	16.73 %	5.77 %
Large net wealth variation	334	0.32 %	0.11 %
Ambiguous labour market transfers	2 820	2.67 %	0.92 %
Ambiguous value of the *nw* variable	9437	8.93 %	3.08 %
Total	105 693	100.00 %	34.47 %
Observations left in the sample	200 921		

Table 20 Fractions of workers fulfilling individual AFP eligibility criteria by age.

Age	Fraction of employed
50-59	0.00 %
60	40.54 %
61	48.38 %
62	64.61 %
63	65.68 %
64	60.55 %
65	53.62 %
66	46.93 %
67	54.57 %
68	30.84 %
69	8.65 %
70	4.13 %

Table 21. Model for employment earnings in period t.

	Dependent variable	wage income					
Estimated on (filter)	Labour market transition	From/to active labour market state					
	Age, other	<60		60-67		≥67	
Applied to (filter)	Labour market transition	From/to active labour market state					
	Age, other	<60		60-67		≥67	
Variable		estim.	st.err.	estim.	st.err.	estim.	st.err.
Constant term		33.046	0.508	-28.438	0.903	-82.030	7.764
Gender		16.474	0.151	13.339	0.207		
Time index (age-50)		-1.250	0.024				
Time index squared				-0.166	0.002		
Number of high wage years (*nw*)		-2.080	0.052	-4.591	0.064		
Aggregate wage (*aw*)		0.229	0.004	1.580	0.012	2.240	0.108
Aggregate wage squared (aw^2)		0.001	0.000	-0.007	0.000	-0.009	0.000
Aggregate wage to the third (aw^3)				9.9E-06	0.000	1.5E-05	0.000
Cross effect *nw***aw*		0.042	0.000	0.060	0.000	0.053	0.001
Current period unemployment		(reference)		(reference)			
Current period non-AFP employment		48.070	0.253	46.185	0.363		
Curr. partial DI + non-AFP employment		-18.853	0.383	-14.388	0.446		
Current period AFP employment		51.332	0.247	56.137	0.356		
Curr. partial DI + AFP employment		-8.728	0.376	-7.919	0.448		
F-test		.		.		5116.55	
p-value for F-test		0.0000		0.0000		0.0000	
Number of observations		914 839		703 107		11 296	
R-square		72.36 %		66.63 %		64.45 %	
Estimated standard error of residuals		50.767		63.028		80.212	

Table 22. Model for pension incomes (AFP and NIS) in period t.

	Dependent variable	Pension income (AFP and NIS)			
Estimated on (filter)	Labour market transition	From/to pension state			
	Age, other	afp-67		≥67	
Applied to (filter)	Labour market transition	From any to pension state			
	Age, other	afp-67		≥67	
Variable		estim.	st.err.	estim.	st.err.
Constant term		161.099	12.571	-1692.08	66.210
AFP age		-1.901	0.198	-7.880	0.301
Gender		-6.244	0.395	-15.981	0.463
Time index (age-50)				242.880	7.053
Time index squared				-6.442	0.190
Number of high wage years (*nw*)				-4.964	0.080
Aggregate wage (*aw*)		0.532	0.011	0.534	0.009
Aggregate wage squared (aw^2)		-0.001	0.000	-0.001	0.000
F-test		6293.00		4464.25	
p-value for F-test		0.0000		0.0000	
Number of observations		53 212		108 727	
R-square		32.12 %		22.33 %	
Estimated standard error of residuals		31.133		55.415	

Table 23. Model for pension incomes (full and part-time disability) in period t.

	Dependent variable	Pension income (disability)			
Estimated on (filter)	Labour market transition	From/to full DI		From/to part DI	
Applied to (filter)	Labour market transition	To full DI		To part DI	
Variable		estim.	st.err.	estim.	st.err.
Constant term		375.909	4.341	21.978	0.673
AFP age		-5.288	0.069		
Gender		-6.690	0.119	-8.595	0.197
Time index (age-50)				0.617	0.024
Number of high wage years (*nw*)				-2.221	0.048
Aggregate wage (*aw*)		0.403	0.003	0.243	0.002
Aggregate wage squared (aw^2)		-1.18E-4	6.15E-6		
Cross effect *nw***aw*				-0.010	0.0005
Previous period unemployment				(reference)	
Prev. non-AFP employment				7.802	0.569
Prev. partial DI + non-AFP employment				16.200	0.536
Prev. AFP employment				20.509	0.582
Prev. partial DI + AFP employment				16.365	0.564
Curr. partial DI + non-AFP employment				(reference)	
Curr. partial DI + AFP employment				-7.979	0.279
F-test		84202.17		7667.68	
p-value for F-test		0.0000		0.0000	
Number of observations		216 503		151 419	
R-square		60.87 %		35.74 %	
Estimated standard error of residuals		21.466		30.311	

Table 24. Model for additional incomes in period t.

Dependent variable	Additional income			
Model	Logit for positive values of additional income		Ordinary regression	
Estimated on (filter)	All observations		Observations with positive add. income	
Applied to (filter)	All observations		Observations with logit prediction >0.5	
Variable	estim.	st.err.	estim.	st.err.
Constant term	-0.747	0.014	25.993	0.389
Gender	0.325	0.006	12.928	0.112
Time index (age-50)	0.167	0.0007	1.408	0.056
Time index squared			-0.102	0.003
Spouse indicator (*sp*)	-1.081	0.005	-5.121	0.089
Number of high wage years (*nw*)	-0.093	0.001	-1.017	0.012
Aggregate wage (*aw*)	0.0004	0.00003	0.118	0.002
Aggregate wage squared (aw^2)			3.33E-4	4.53E-6
Current period OLM		(reference)	63.230	0.193
Current period pension retirement	0.346	0.009	-17.351	0.185
Current period full disability	3.157	0.014	-21.772	0.144
Current period unemployment	-1.042	0.012	9.563	0.296
Current period non-AFP employment	-1.337	0.009		(reference)
Curr. partial DI + non-AFP employment	3.257	0.019	-29.762	0.175
Current period AFP employment	-1.538	0.009		(reference)
Curr. partial DI + AFP employment	3.728	0.021	-24.644	0.171
F-test				44437.03
p-value for F-test				0.0000
Number of observations		2 262 566		886 269
R-square (pseudo for logit)		50.88 %		39.46 %
Estimated standard error of residuals				38.659

Table 25. Model for spouse incomes in period t.

Dependent variable	Spouse income			
Estimated on (filter)	Spouse exists *gender*=0		Spouse exists *gender*=1	
Applied to (filter)	Spouse exists *gender*=0		Spouse exists *gender*=1	
Variable	estim.	st.err.	estim.	st.err.
Constant term	-502.222	9.063	664.095	15.916
AFP age	9.635	0.145	-8.726	0.256
Time index (age-50)	-8.608	0.086	-17.124	0.165
Time index squared	0.069	0.005	1.080	0.008
Number of high wage years (*nw*)			-0.601	0.046
Aggregate wage (*aw*)	0.365	0.007	0.245	0.003
Aggregate wage squared (aw^2)	-0.001	0.000		
Current period OLM	(reference)			
Current period pension retirement	21.226	0.498		
Current period full disability	10.907	0.483		
Current period unemployment	9.828	0.628		
Current period non-AFP employment	9.071	0.450		
Curr. partial DI + non-AFP employment	11.589	0.619		
Current period AFP employment	18.651	0.443		
Curr. partial DI + AFP employment	18.698	0.710		
F-test	10892.77		8595.85	
p-value for F-test	0.0000		0.0000	
Number of observations	1 076 041		433340	
R-square (pseudo for logit)	10.83 %		9.02 %	
Estimated standard error of residuals	83.960		110.555	

Table 26. Listing of the C function for transition probability calculation.

```
double pr (int age,int ps,int h,int m,int e,int sp,int nw,int awi,double aw,int cs,
           int fh, int fm, int fe, int fsp, int fnw, double faw)
{   /*Probability of combination of next period (fh,fm,fe,fsp,fnw)
      with current age,ps,h,m,e,sp,nw,awi,aw,cs, global arrays h_prtr and m_prtr, coef_trpr array */
    double pr=1.0, res;
    if (cs==0 || cs==2)
    {   //special cases with most combinations of future state vars infeasible
        if (fm!=0 || fe!=0)
        { pr=0; }
        else
        {   //Health
            pr=pr*h_prtr[fh*nh+h];
        }
    }
    else if (cs==1)
    {   if (fm!=0)
        { pr=0; }
        else
        {   //Health
            pr=pr*h_prtr[fh*nh+h];
            //zero in all infeasible combinations!
        }
    }
    else
    {   //regular case
        //for s=5,7 bad health has zero pr by setup of the transition matrix
        //Health
        pr=pr*h_prtr[fh*nh+h];
        //Matching (dependent on same period health)
        if (fh==0) pr=pr*m_prtr_h0[fm*nm+m];
        if (fh==1) pr=pr*m_prtr_h1[fm*nm+m];
        if (fh==2 && fm!=0) pr=0; //no matching for bad bad health
        if (fh==2 && fm==0) pr=pr;
        //AFP eligibility
        if (age+1[afp || age+1]=retage1)
        {   // e===0 for all for these ages
            if (fe==0) pr=pr;
            if (fe==1) pr=0;
        }
        if (age+1==afp && afp<retage1)
        {   //group 1
            if (cs<=5 || nw<2 || aw<74)
            {   //perfect prediction: fe=0
                if (fe==0) pr=pr;
                if (fe==1) pr=0;
            }
            else
            {   //logit model from coef_trpr (column 1)
                res=coef_trpr[0]+
                    coef_trpr[3]*(nw==9?1:0) + coef_trpr[4]*(nw==10?1:0);
                if (fe==0) pr=pr/(1+exp(res));
                if (fe==1) pr=pr*exp(res)/(1+exp(res));
            }
        }
        if (age+1>afp && age+1<retage1)
        {   //group 2
            if (cs<=5 || nw<2)
            {   //perfect prediction: fe=0
                if (fe==0) pr=pr;
                if (fe==1) pr=0;
            }
            else
            {
                if (e==1)
                {   //perfect prediction: fe=1
                    if (fe==0) pr=0;
                    if (fe==1) pr=pr;
                }
                else
                {   //logit model from coef_trpr (column 2)
                    res=coef_trpr[5+0] + coef_trpr[5+1]*afp + coef_trpr[5+2]*gender + coef_trpr[5+3]*aw + coef_trpr[5+4]*nw;
                    if (fe==0) pr=pr/(1+exp(res));
                    if (fe==1) pr=pr*exp(res)/(1+exp(res));
                }
            }
        }
    }
    //Spouse survival
    res=surpr(age+1,2);
    if (sp==1 && fsp==1) pr=pr*res;
    if (sp==1 && fsp==0) pr=pr*(1-res);
    if (sp==0 && fsp==1) pr=0;
    if (sp==0 && fsp==0) pr=pr;
    //years with wage over 1G : nw
    if (fnw!=0 && fnw!=nw+1 && !(nw==nnw-1 && fnw==nnw-1))
    {   //wrong value of fnw
        pr=0;
    }
    else if (nnw-1==0)
    {   //trivial case with nnw=1 (simplified test runs) pr is called twice with fnw=0..
        pr=pr/2;
    }
    else
    {   //logit model from coef_trpr (column 3)
        res=coef_trpr[10+0]+
            coef_trpr[10+1]*(cs==3||cs==4||cs==6?1:0) + coef_trpr[10+2]*(cs==5||cs==7?1:0)
          + coef_trpr[10+3]*(cs==0||cs==2?1:0) + oef_trpr[10+4]*(nw==0?1:0);
        if (fnw==nw+1 || fnw==nnw-1) pr=pr*exp(res)/(1+exp(res));
        if (fnw==0) pr=pr/(1+exp(res));
    }
    if (v_tree(age+1,fh,fm,fe,cs)==0) { pr=0; } //infeasible next year combination
    return pr;
}
```

Table 27. The tax function.

	Model	OLS	
Number of observation		160 160	
Goodness of fit		98.34%	
Estimated standard error of residuals		15.992	
Explanatory variables		Coeff.	Std. Err.
$I(t, d_t, s_t)$	Household income	0.423	0.000
sp_t	Type of the household (0 for single, 1 for full)	-21.951	0.080
$ps_{t+1} \in \{1, 2\}$	Curr. state: pension, disability	-18.718	0.103
$ps_{t+1} \geq 4$	Curr. state: working, partial disability	-9.250	0.103
-	Constant term	-24.501	0.113
All coefficients significant at 0.0005 level			

Table 28. Simulated distributions of states on the labour market by age (calibration).

State	50	51	52	53	54	55	56	57	58	59
0	0.0000	0.0380	0.0401	0.0422	0.0485	0.0549	0.0611	0.0620	0.0717	0.0725
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0060	0.0100	0.0121	0.0183	0.0244	0.0372	0.0422	0.0576
3	0.0800	0.0160	0.0160	0.0161	0.0222	0.0264	0.0183	0.0248	0.0295	0.0192
4	0.3380	0.3720	0.3367	0.3173	0.2949	0.2663	0.2648	0.2603	0.2553	0.2473
5	0.0200	0.0320	0.0341	0.0341	0.0545	0.0508	0.0530	0.0682	0.0696	0.0746
6	0.5520	0.5080	0.5090	0.5080	0.4828	0.4736	0.4542	0.4215	0.3966	0.3923
7	0.0100	0.0340	0.0581	0.0723	0.0848	0.1098	0.1242	0.1260	0.1350	0.1365
60	61	62	63	64	65	66	67	68	69	70
0.0941	0.0962	0.1002	0.1063	0.1086	0.1384	0.1709	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0706	0.1280	0.1869	0.2402	0.2269	0.8896	0.9180	0.9362	1.0000
0.0744	0.0850	0.1025	0.1304	0.1515	0.1854	0.2801	0.0000	0.0000	0.0000	0.0000
0.0219	0.0336	0.0228	0.0169	0.0126	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2473	0.2327	0.2164	0.1908	0.1515	0.1123	0.0756	0.0478	0.0252	0.0284	0.0000
0.0722	0.0828	0.0752	0.0870	0.0833	0.0601	0.0364	0.0000	0.0000	0.0000	0.0000
0.3414	0.3289	0.2597	0.1908	0.1566	0.1149	0.1148	0.0627	0.0568	0.0355	0.0000
0.1488	0.1409	0.1526	0.1498	0.1490	0.1488	0.0952	0.0000	0.0000	0.0000	0.0000

A.3. Figures

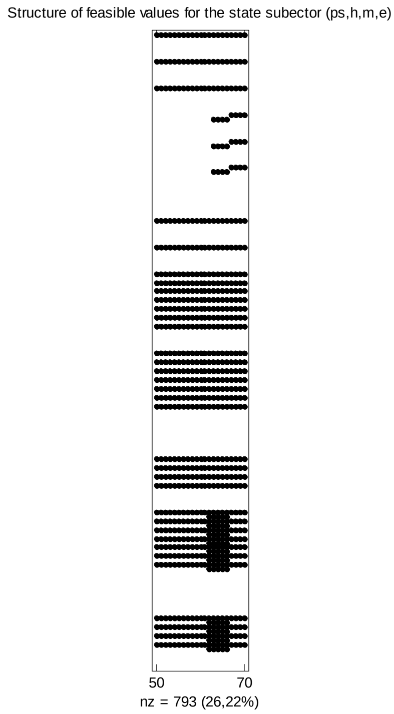

Figure 11. Structure of the feasible values of the state subvector (ps_t, h_t, m_t, e_t) (with AFP age of 62 and normal retirement age of 67 years).

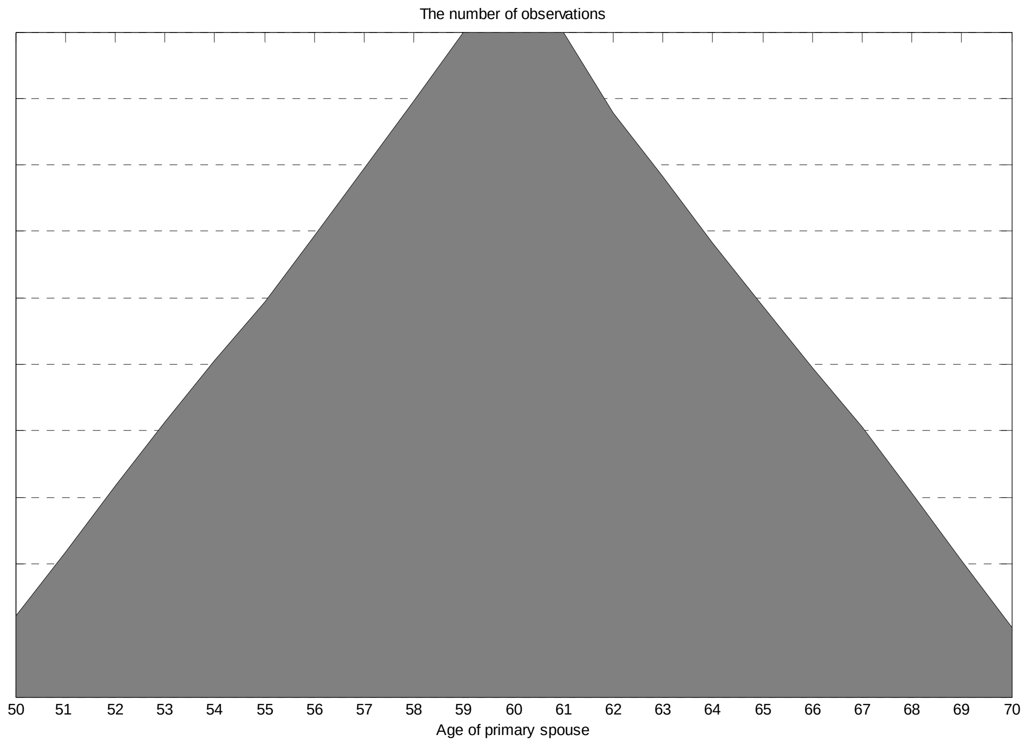

Figure 12. Shape of the panel of observations.

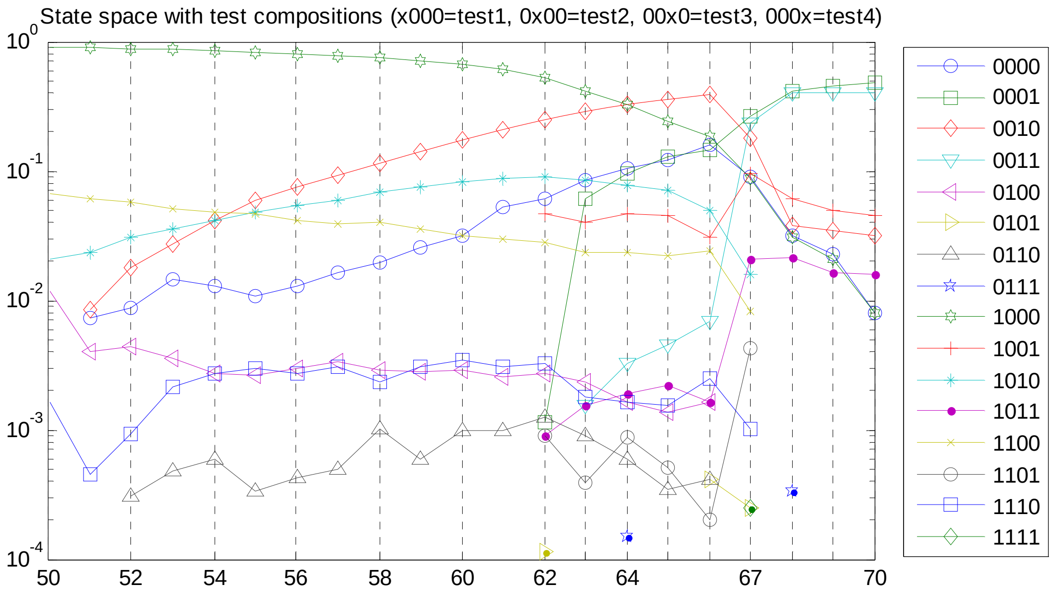

Figure 13. Observed combinations of labour market states tests (see explanations in the text).

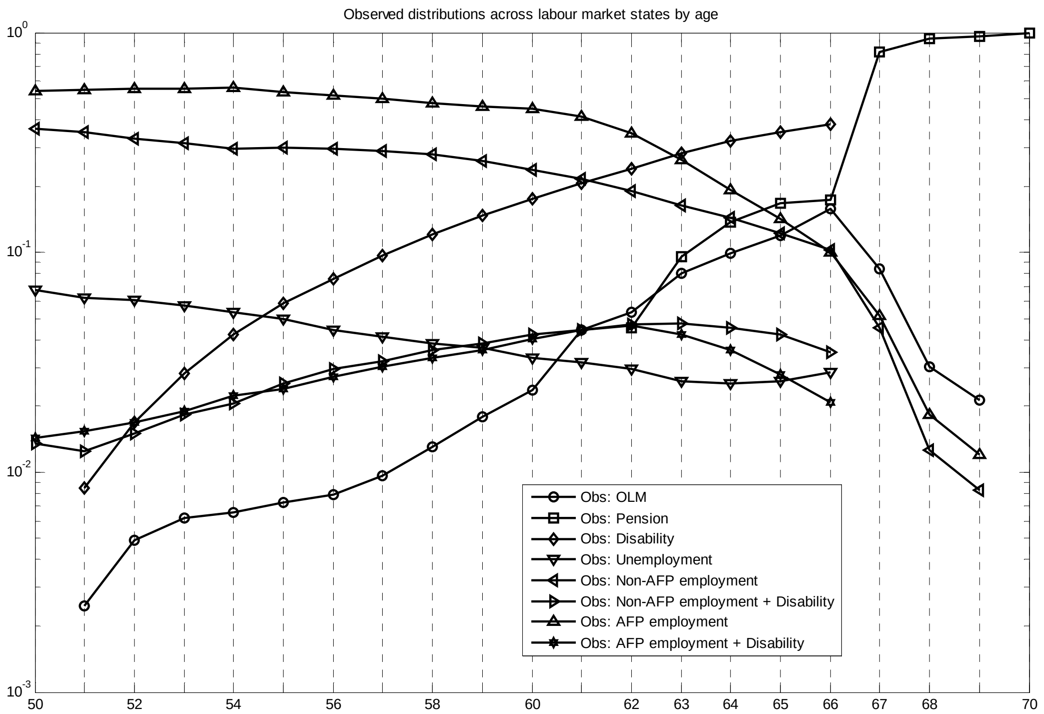

Figure 14. Observed fractions of states on the labour market by age.

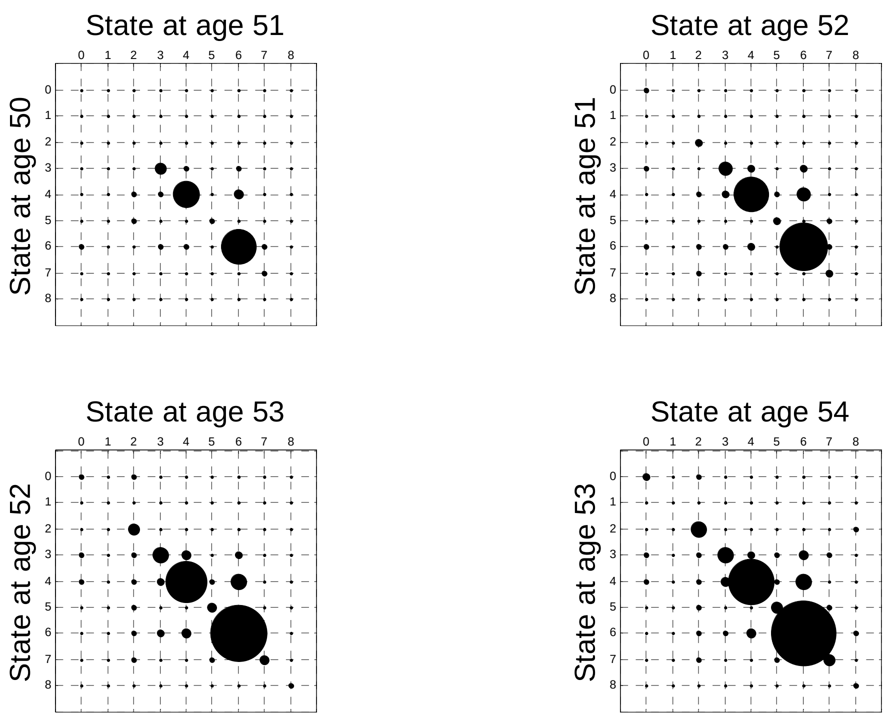

Figure 15. Observed transitions among states on the labour market (1 of 5)

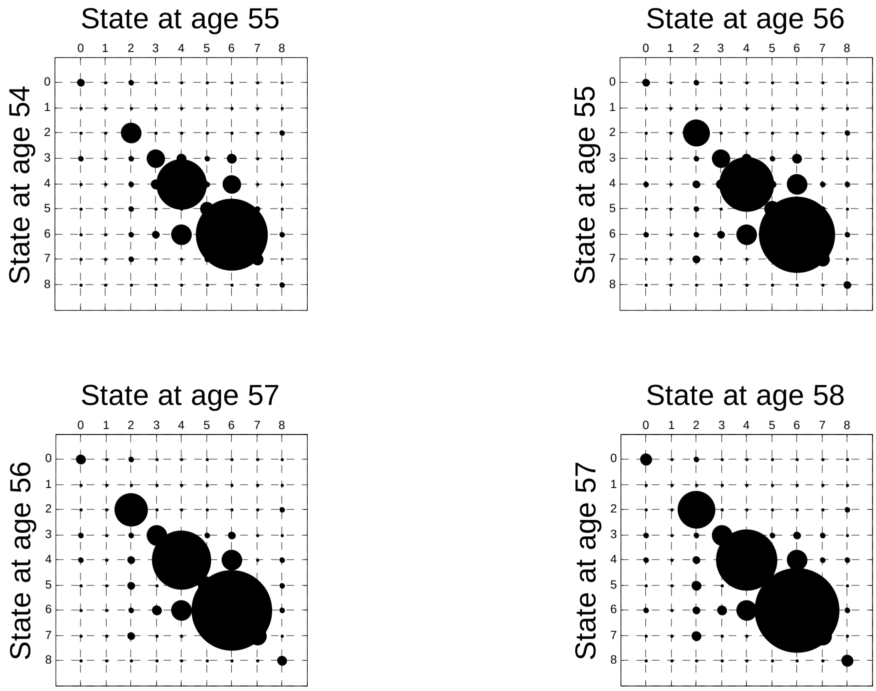

Figure 16. Observed transitions among states on the labour market (2 of 5)

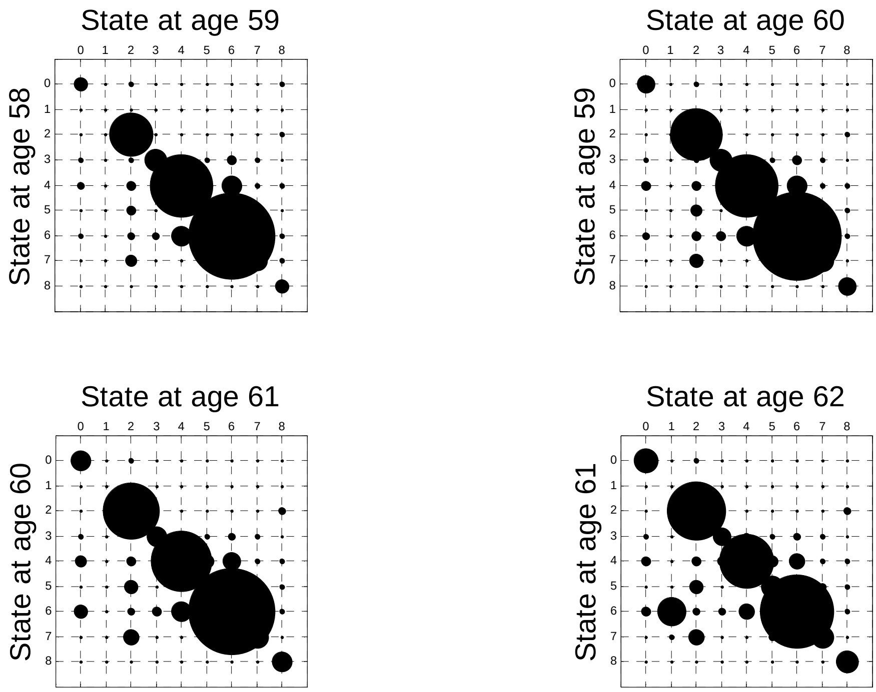

Figure 17. Observed transitions among states on the labour market (3 of 5)

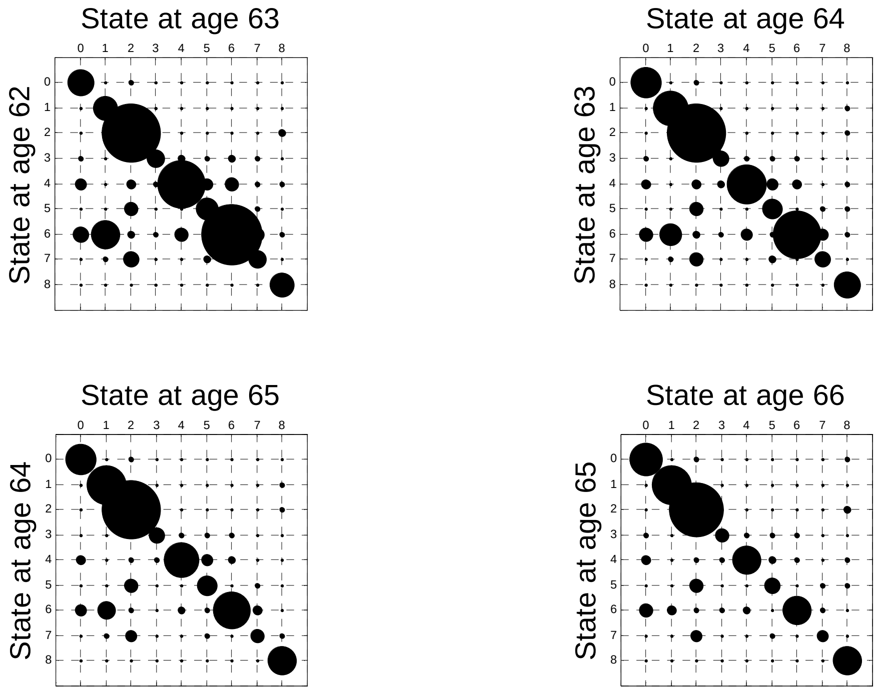

Figure 18. Observed transitions among states on the labour market (4 of 5)

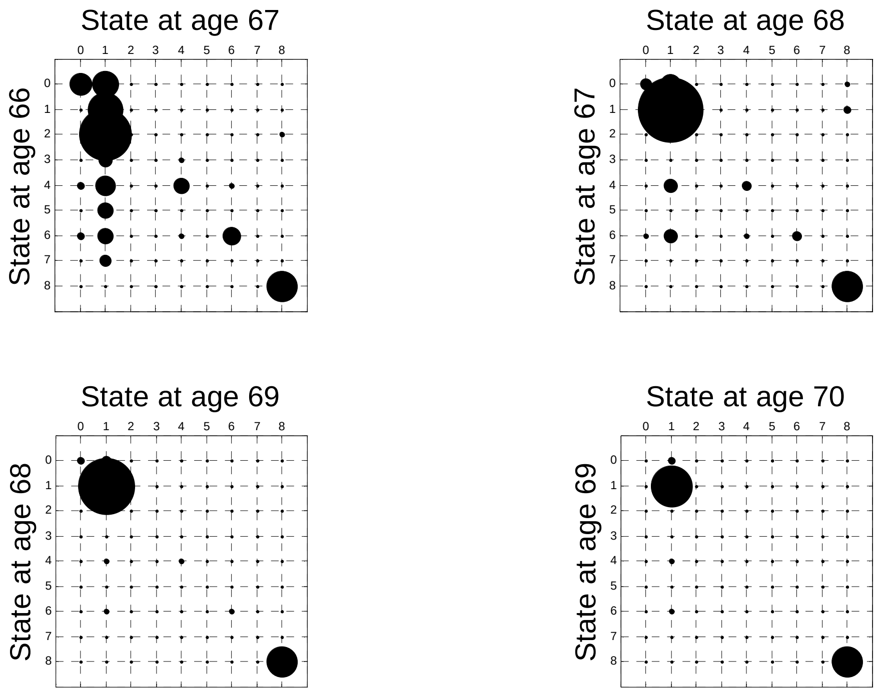

Figure 19. Observed transitions among states on the labour market (5 of 5)

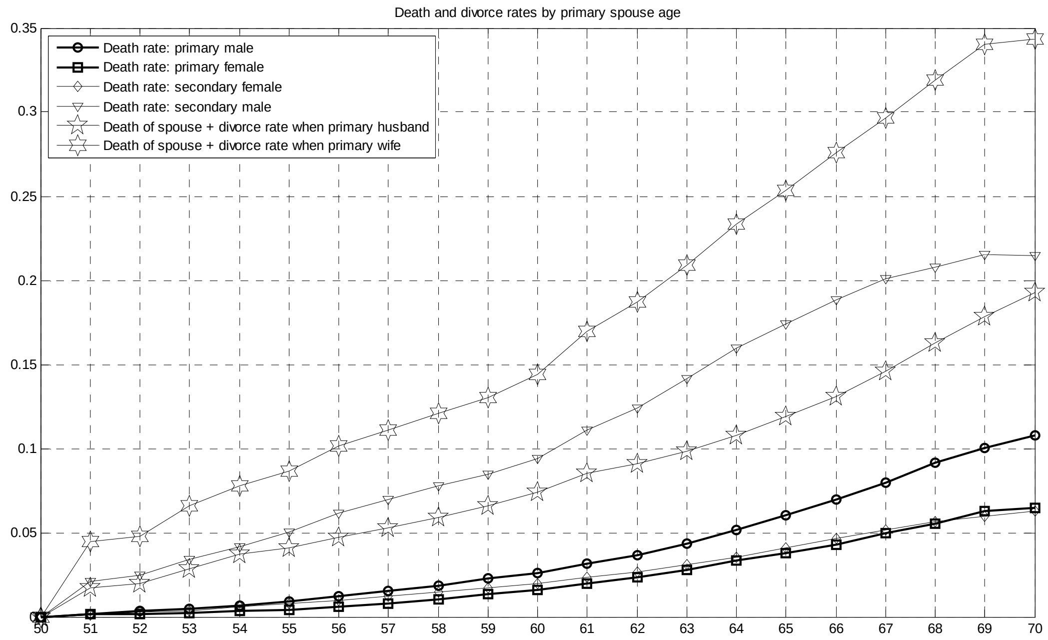

Figure 20. Graphical representation of death tables and divorce tables.

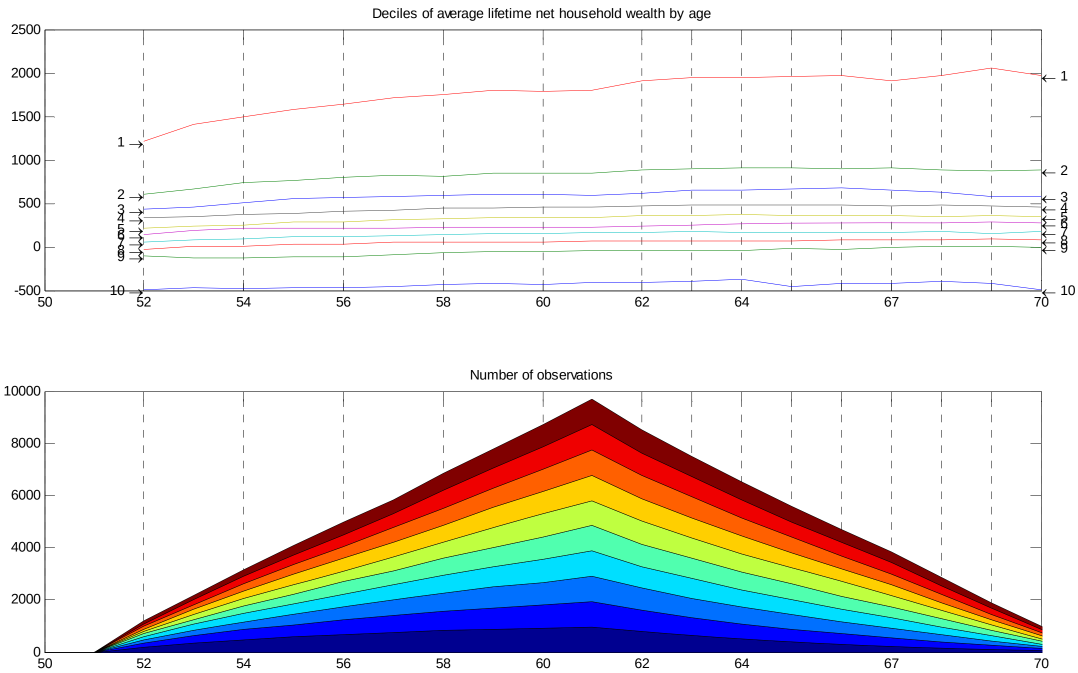

Figure 21. Net wealth dynamics by age. Number of observations of wealth.

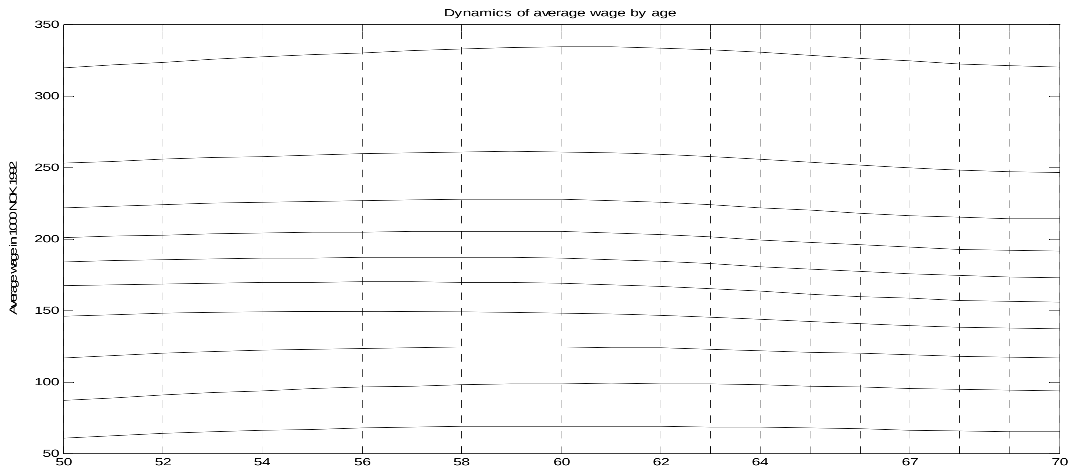

Figure 22. Average wage dynamics by age.

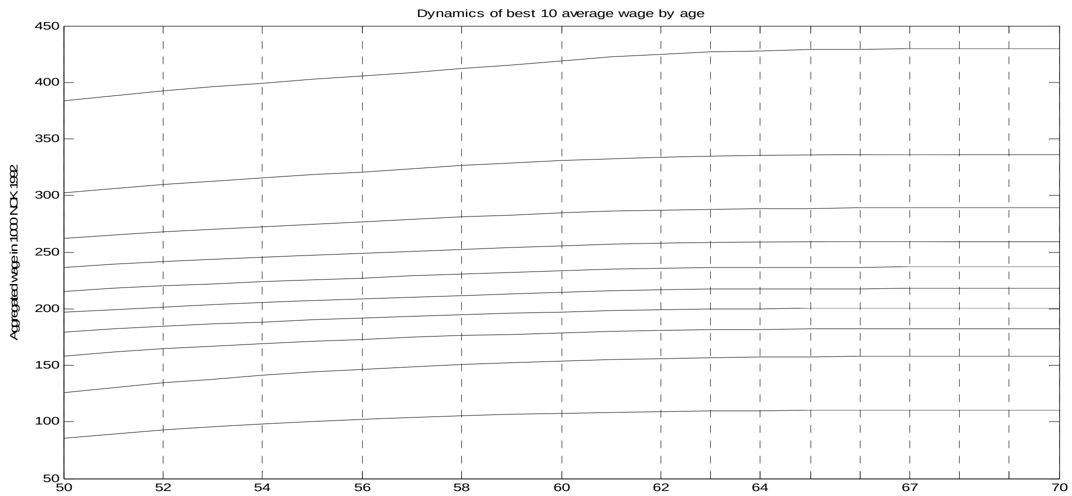

Figure 23. Best 10 average wage dynamics by age.

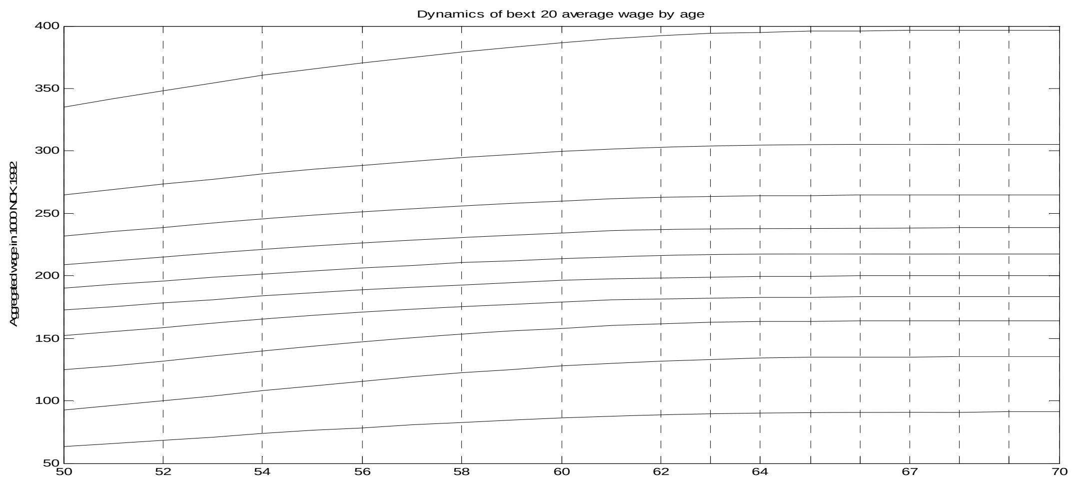

Figure 24. Best 20 average wage dynamics by age.

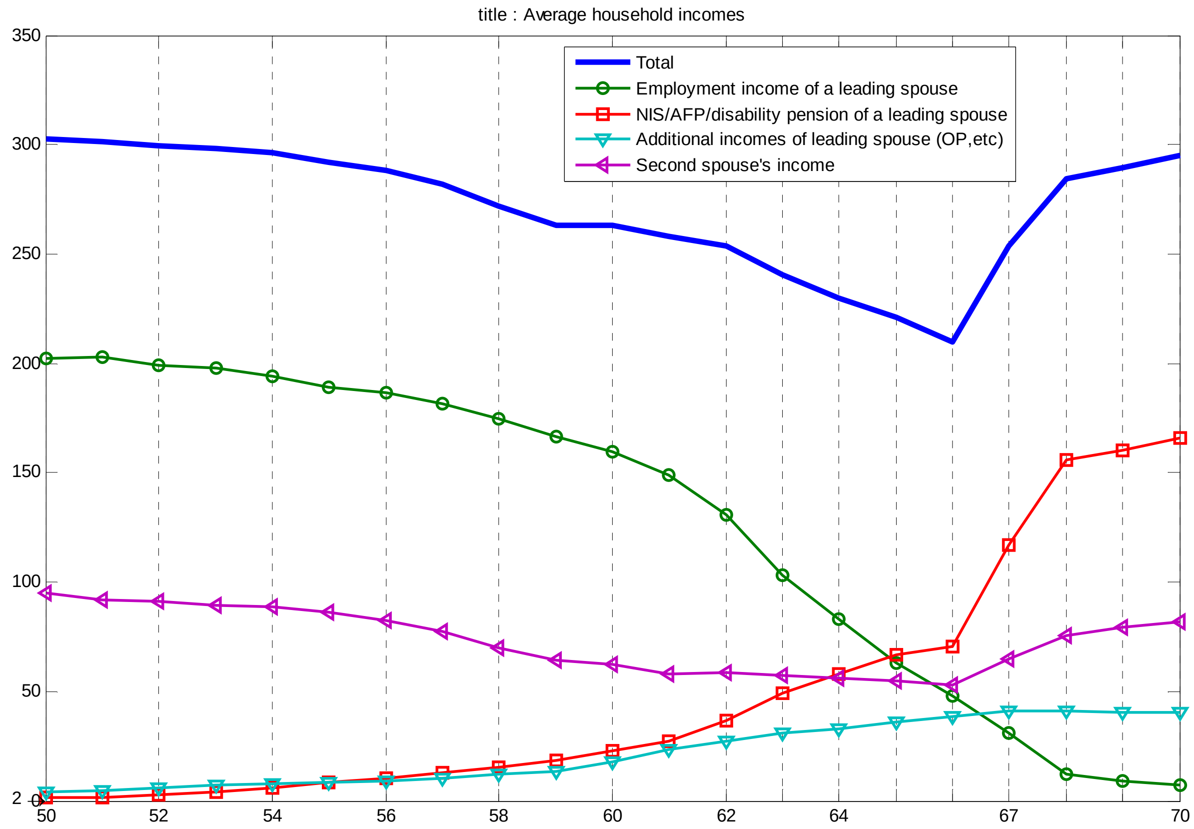

Figure 25. Average household incomes by age (thousand NOK in 1992 prices).

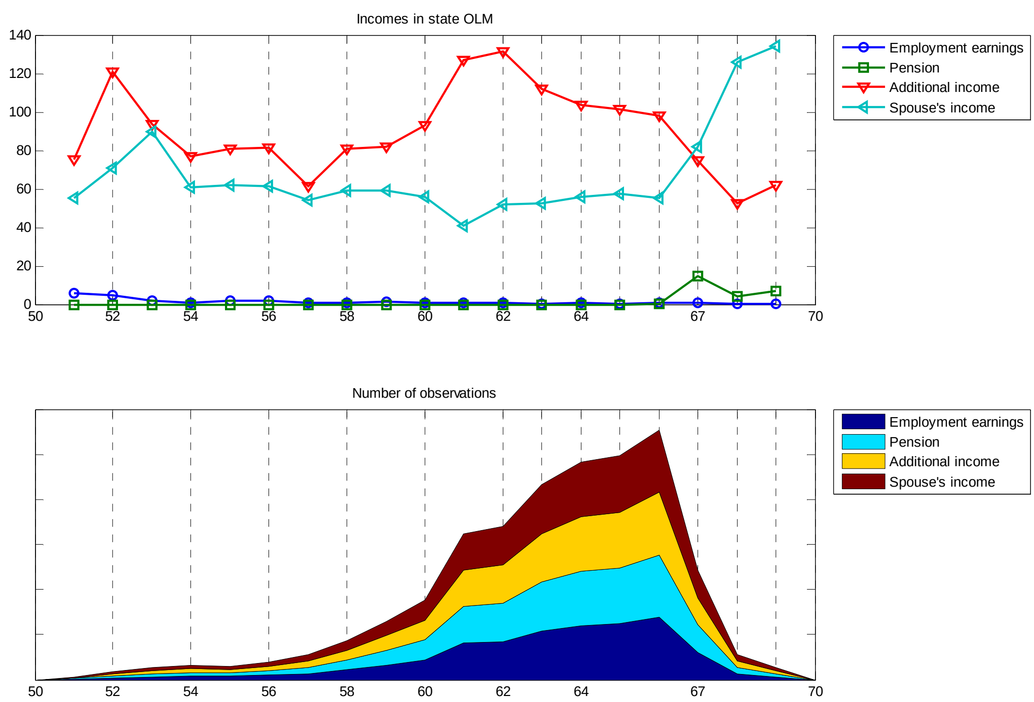

Figure 26. Incomes in OLM state.

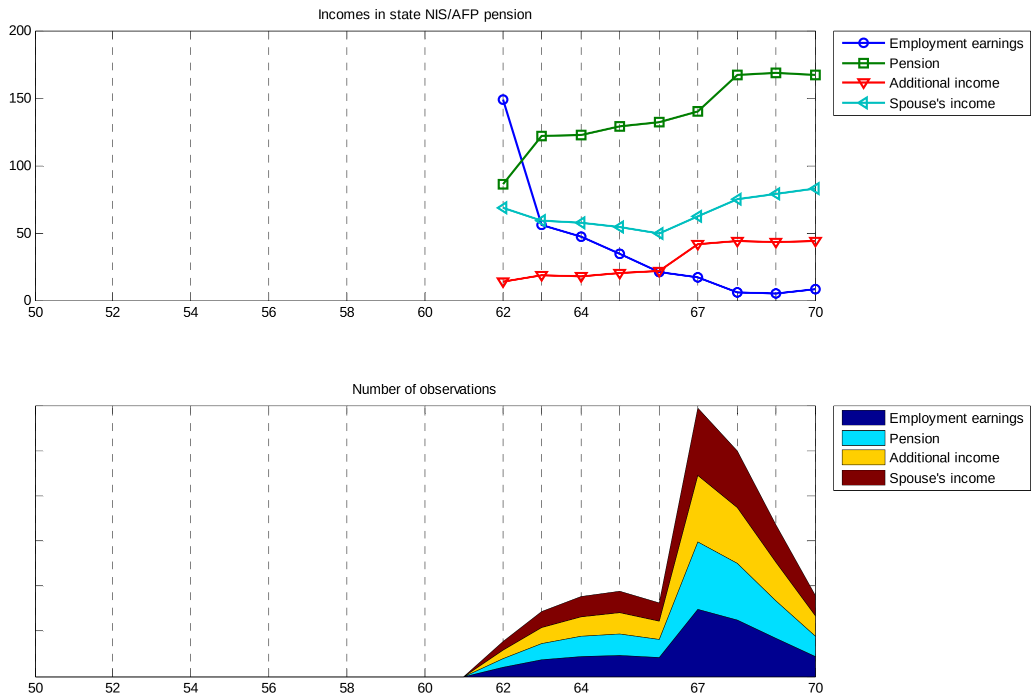

Figure 27. Incomes in pension state.

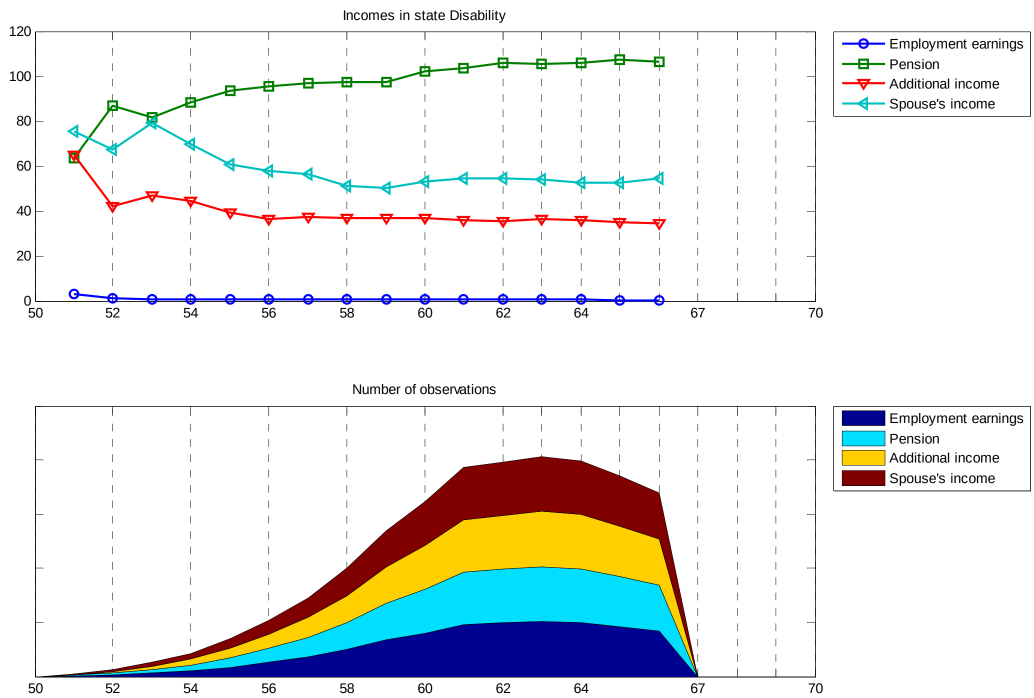

Figure 28. Incomes in disability state.

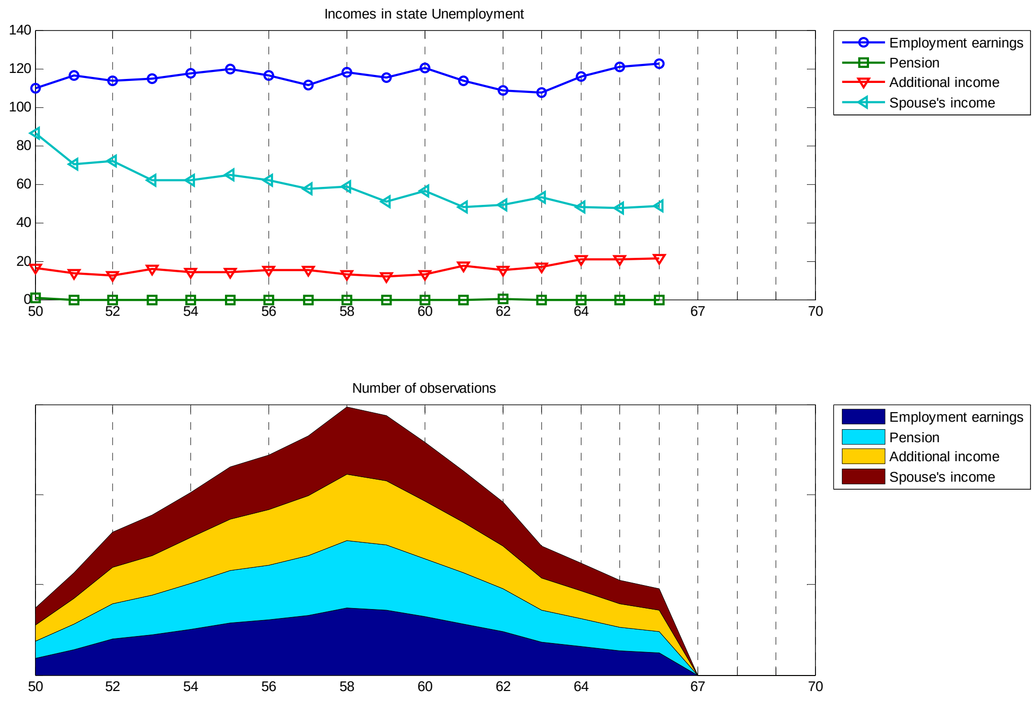

Figure 29. Incomes in unemployment state.

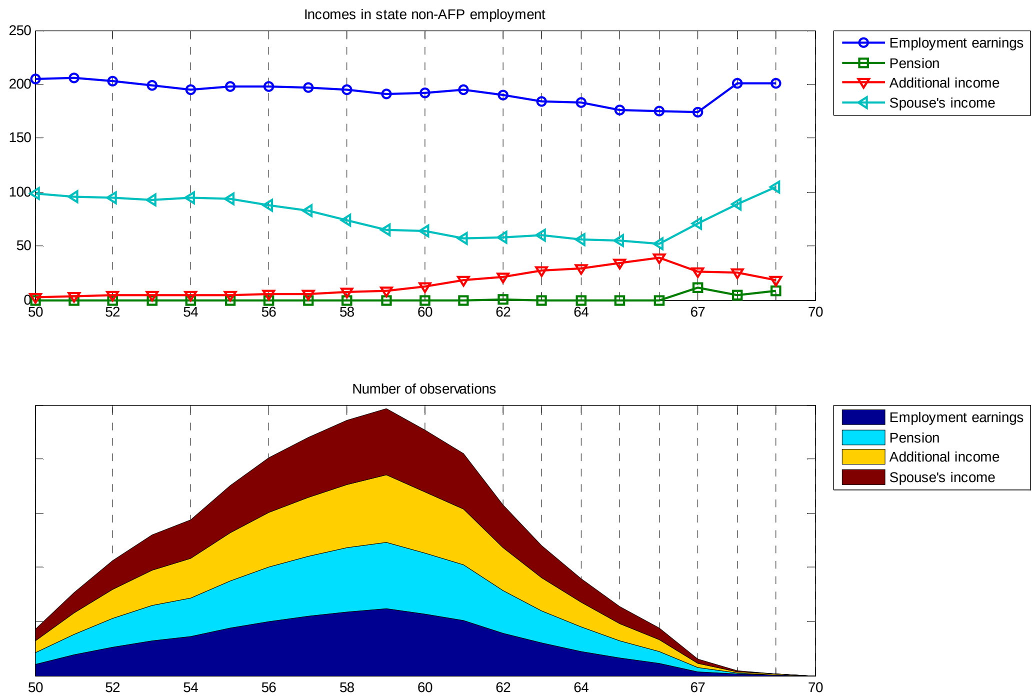

Figure 30. Incomes in non-AFP employment state.

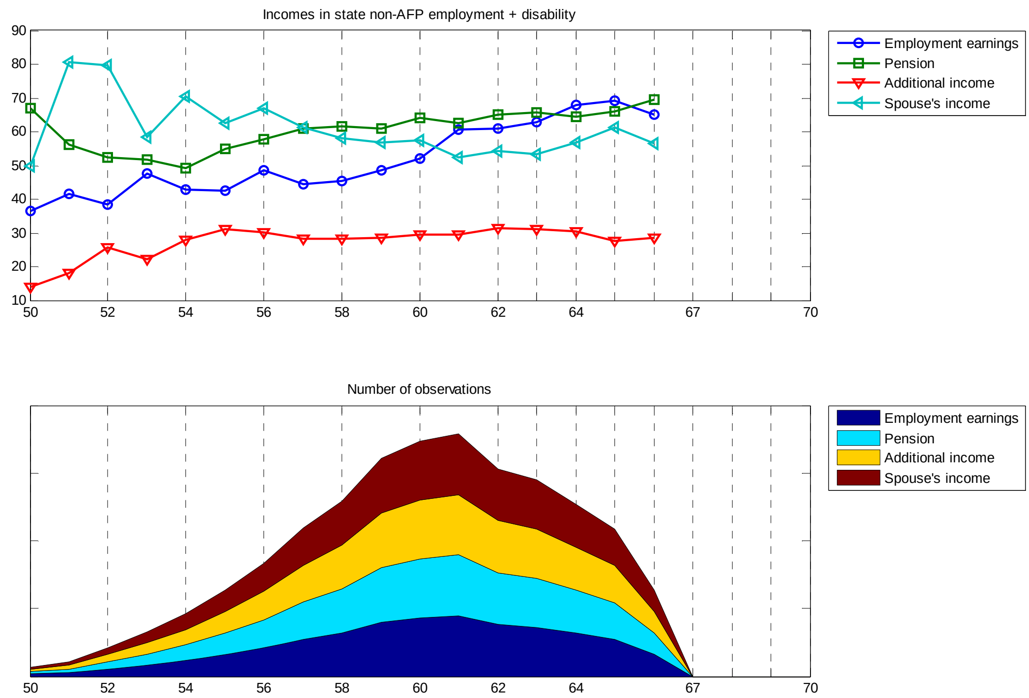

Figure 31. Incomes in non-AFP employment + disability state.

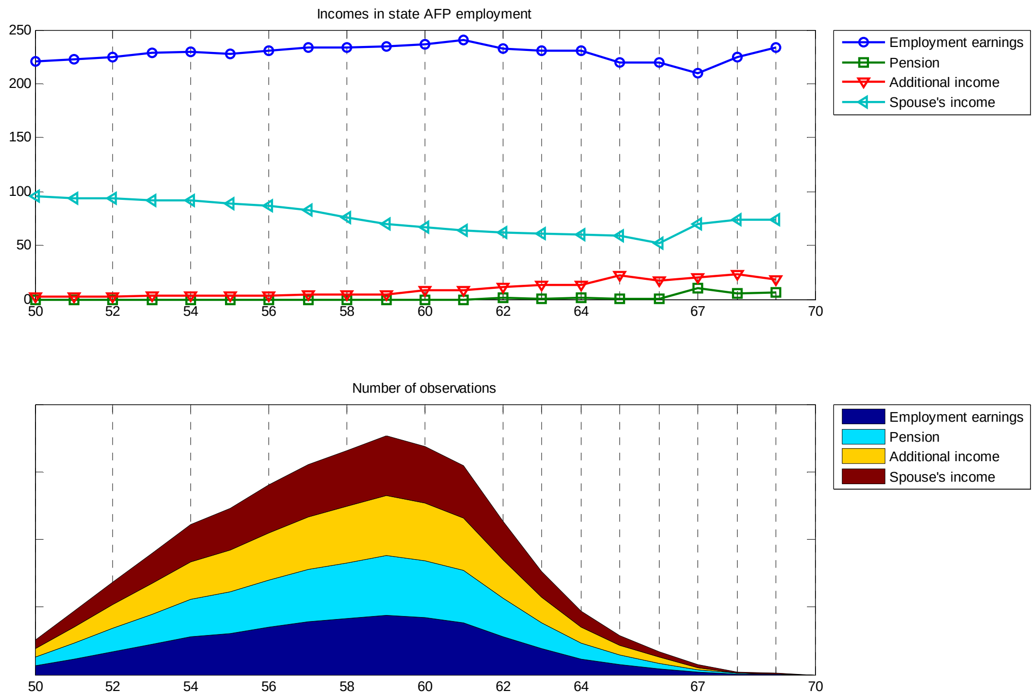

Figure 32. Incomes in AFP employment state.

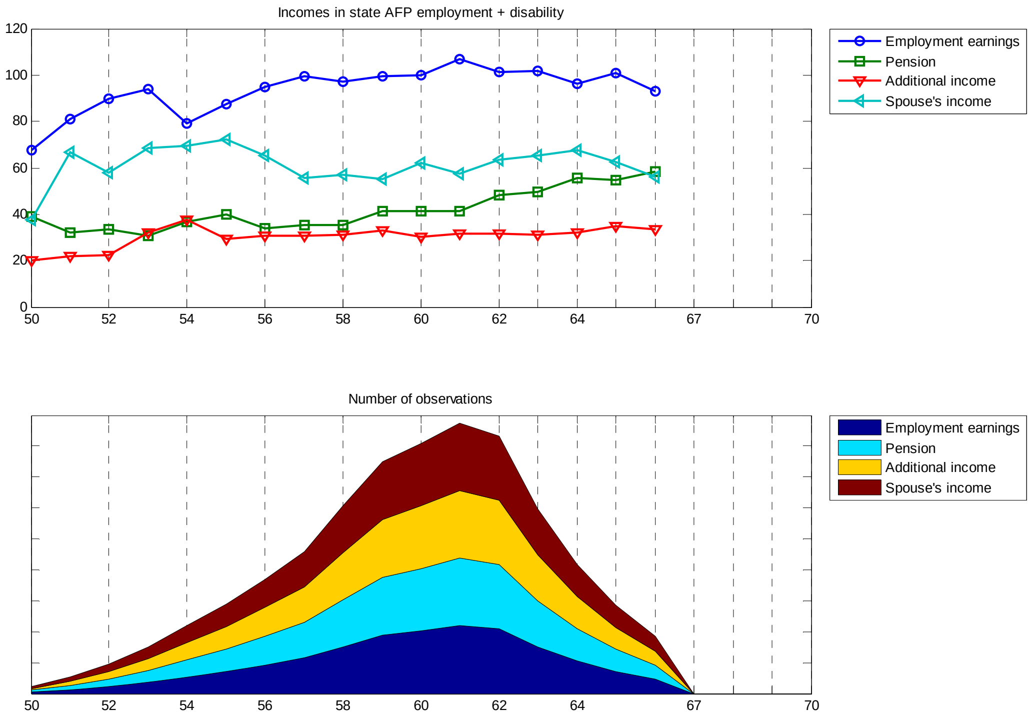

Figure 33. Incomes in AFP employment + disability state.

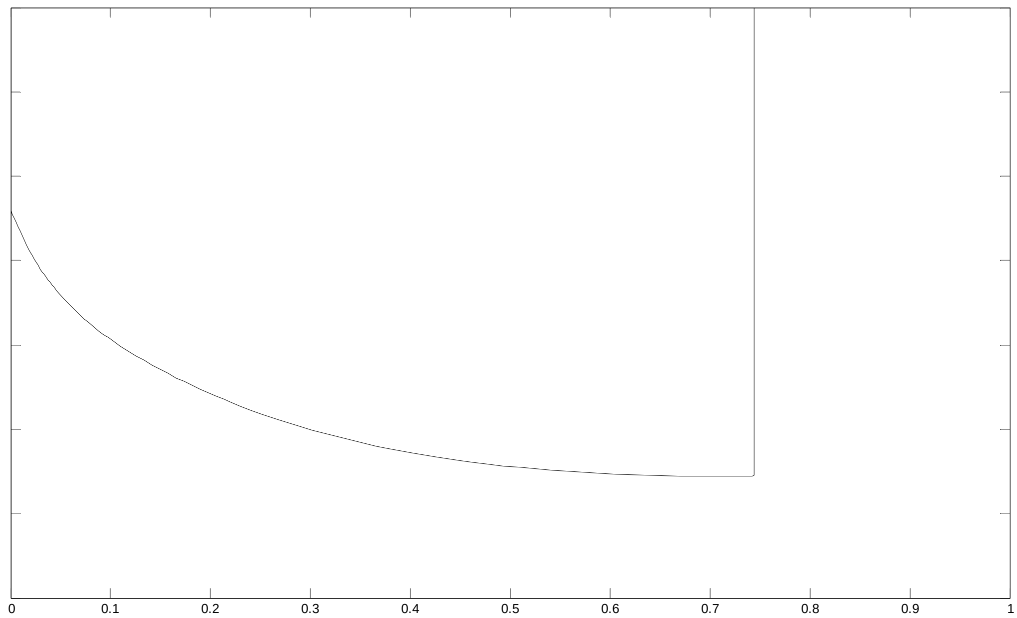

Figure 34. Typical likelihood as a function of the last parameter in the termination function.

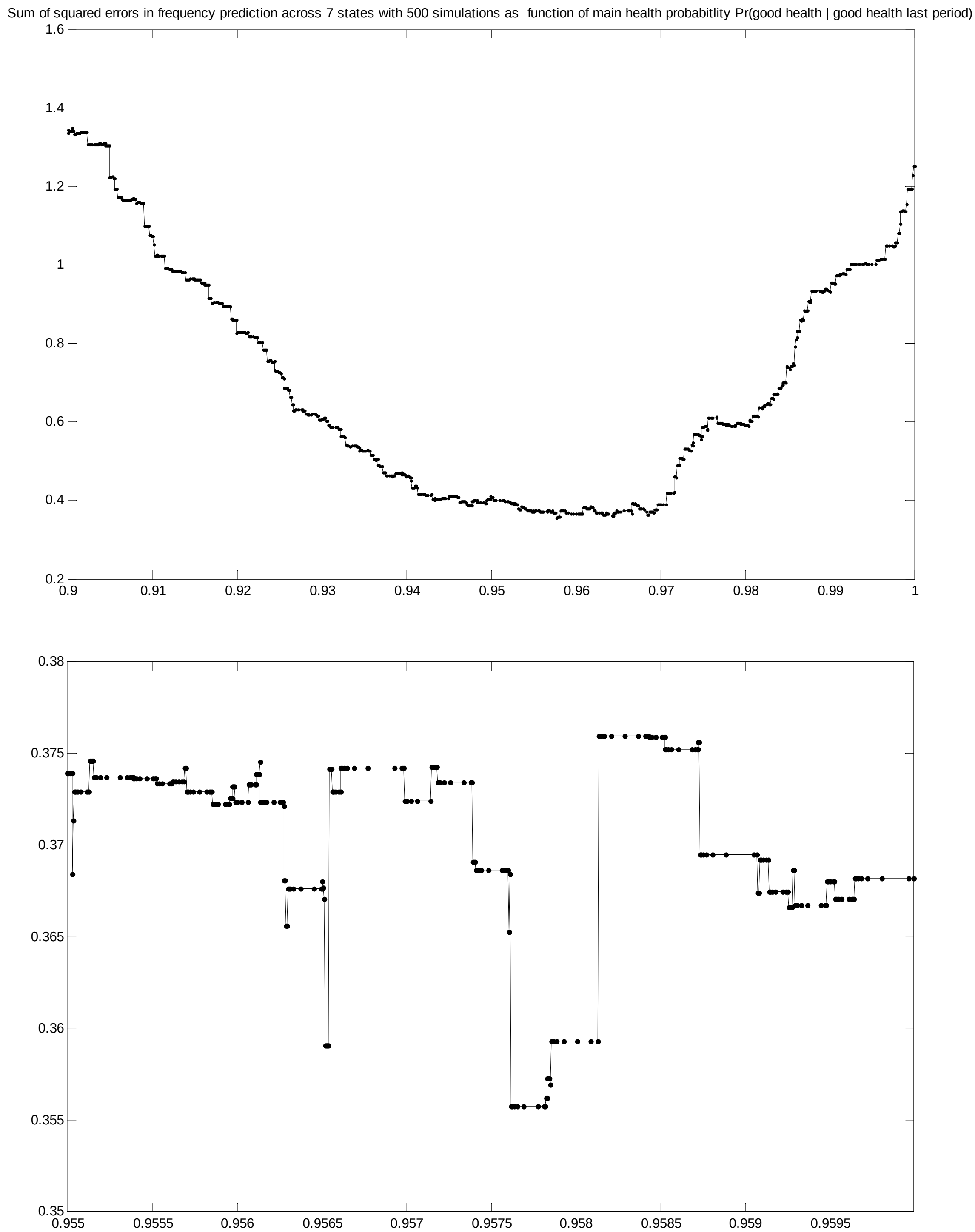

Figure 35. Sum of squared errors in prediction of frequencies of sample in 7 states and 20 periods with 1000 simulated decision making agents as a function of main health probability $\pi_{00}^{(h)}$.

References

1. Allison, R.A. and J.E. Foster (2004). "Measuring health inequality using qualitative data." Journal of Health Economics **23**(3): 505-524.
2. Au, D.W.H., T.F. Crossley and M. Schellhorn (2005). "The effect of health changes and long-term health on the work activity of older Canadians." Health Economics **14**(10): 999-1018.
3. Autor, D.H. and M.G. Duggan (2006). "The growth in the social security disability rolls: A fiscal crisis unfolding." Journal of Economic Perspectives **20**(3): 71-96.
4. Bellman, R.a.S.D. (1962). Applied dynamic programming. Princeton, New Jorsey, Princeton University Press.
5. Benitez-Silva, H., M. Buchinsky, H.M. Chan, S. Cheidvasser and J. Rust (2004). "How large is the bias in self-reported disability?" Journal of Applied Econometrics **19**(6): 649-670.
6. Berkovec, J. and S. Stern (1991). "Job Exit Behavior of Older Men." Econometrica **59**(1): 189-210.
7. Blau, D.M. (1994). "Labor-Force Dynamics of Older Men." Econometrica **62**(1): 117-156.
8. Blau, D.M. (1997). "Social security and the labor supply of older married couples." Journal of Labour Economics **4 (1997)**: 373-418.
9. Bloom, D.E., D. Canning, M. Moore and National Bureau of Economic Research. (2004). The effect of improvements in health and longevity on optimal retirement and saving. Cambridge, Mass., National Bureau of Economic Research.
10. Borsch-Supan, A.H. and P. Schmidt (2000). "Early retirement in East and West Germany." IZA Reprint Series A **69/2000**: 83-102.
11. Bound, J. (1989). "The Health and Earnings of Rejected Disability Insurance Applicants." American Economic Review **79**(3): 482-503.
12. Bound, J. (1991). "Self-Reported Versus Objective Measures of Health in Retirement Models." Journal of Human Resources **26**(1): 106-138.
13. Bound, J. (1998). The dynamic effects of health on the labor force transitions of older workers. Cambridge, MA, National Bureau of Economic Research.
14. Bound, J., M. Schoenbaum and T. Waidmann (1995). "Race and education differences in disability status and labor force attachment in the health and retirement survey." Journal of Human Resources **30**: S227-S267.
15. Bound, J., M. Schoenbaum and T. Waidmann (1996). "Race differences in labor force attachment and disability status." Gerontologist **36**(3): 311-321.
16. Bound, J. and T. Waidmann (1992). "Disability Transfers, Self-Reported Health, and the Labor-Force Attachment of Older Men - Evidence from the Historical Record." Quarterly Journal of Economics **107**(4): 1393-1419.
17. Bourguignon, F. and P.A. Chiappori (1992a). "Collective models of household behavior." European Economic Review **36**: 355-364.
18. Bourguignon, F. and P.A. Chiappori (1992b). "Collective Models of Household Behavior - an Introduction." European Economic Review **36**(2-3): 355-364.
19. Bratberg, E. (1999). "Disability retirement in a welfare state." Scandinavian Journal of Economics **101**(1): 97-114.
20. Bratberg, E., T.H. Holmas and O. Thogersen (2004). "Assessing the effects of an early retirement program." Journal of Population Economics **17**(3): 387-408.
21. Bratberg, E., O.A. Nilsen and K. Vaage (2005). "Intergenerational earnings mobility in Norway: Levels and trends." Scandinavian Journal of Economics **107**(3): 419-435.
22. Burkhauser, R.V., J.S. Butler and G. Gumus (2004). "Dynamic programming model estimates of Social Security Disability Insurance application timing." Journal of Applied Econometrics **19**(6): 671-685.

23. Chan, S., A.H. Stevens and National Bureau of Economic Research. (2003). What you don't know can't help you : pension knowledge and retirement decision making. Cambridge, MA, National Bureau of Economic Research.
24. Coile, C. (2003). Retirement incentives and couples' retirement decisions. Cambridge, MA, National Bureau of Economic Research.
25. Coile, C. (2004). Health shocks and couples' labor supply decisions. Cambridge, Mass., National Bureau of Economic Research.
26. Conti, R.M., E.R. Berndt, R.G. Frank and National Bureau of Economic Research. (2006). Early retirement and public disability insurance applications : exploring the impact of depression. Cambridge, MA, National Bureau of Economic Research.
27. Currie, J., J. Gruber and National Bureau of Economic Research. (1995). Health insurance eligibility, utilization of medical care, and child health. Cambridge, MA, National Bureau of Economic Research.
28. Dagsvik, J.K. (2002). "Discrete choice in continuous time: implications of an intertemporal version of the IIA property." Econometrica **Vol. 70, No. 2, March 2002**: 817-831.
29. Dagsvik, J.K. and Z. Jia (2006). Labor supply as a choice among latent job opportunities : a practical empirical approach. Oslo, Statistics Norway.
30. Dagsvik, J.K. and S. Strøm (1992). Labor supply with non-convex budget sets, hours restriction and non-pecuniary job-attributes. Oslo, Central Bureau of Statistics.
31. Dagsvik, J.K. and S. Strøm (1997). A framework for labor supply analysis in the presence of complicated budget restrictions and qualitative opportunity aspects. Oslo, University of Oslo, Department of Economics.
32. Dagsvik, J.K. and S. Strøm (2006). "Sectoral labour supply, choice restrictions and functional form." Journal of Applied Econometrics **21**(6): 803-826.
33. Dempster, A.P., N.M. Laird and D.B. Rubin (1977). "Maximum Likelihood from Incomplete Data Via Em Algorithm." Journal of the Royal Statistical Society Series B-Methodological **39**(1): 1-38.
34. Disney, R., C. Emmerson and M. Wakefield (2006). "Ill health and retirement in Britain: A panel data-based analysis." Journal of Health Economics **25**(4): 621-649.
35. Dwyer, D.S. and O.S. Mitchell (1999). "Health problems as determinants of retirement: Are self-rated measures endogenous?" Journal of Health Economics **18**(2): 173-193.
36. French, E. (1999). "The effects of health, wealth, and wages on labor supply and retirement behavior." preliminary version **November**.
37. Gilleskie, D.B. (1998). "A dynamic stochastic model of medical care use and work absence." Econometrica **66**(1): 1-45.
38. Gjesdal, S. and E. Bratberg (2003). "Diagnosis and duration of sickness absence as predictors for disability pension: Results from a three-year, multi-register based and prospective study." Scandinavian Journal of Public Health **31**(4): 246-254.
39. Gordon, R.H. and A.S. Blinder (1980). "Market wages, reservation wages and retirement decisions." Journal of Public Economics **14 (1980)**: 277-308.
40. Greene, W.H. (2000). Econometric Analysis. New Jersey, Prentice Hall.
41. Gruber, J. (2000). "Disability insurance benefits and labor supply." Journal of Political Economy **108**(6): 1162-1183.
42. Gruber, J. and B.C. Madrian (2002). Health insurance, labor supply, and job mobility: a critical review of the literature. Cambridge, MA., National Bureau of Economic Research.
43. Gustman, A.L. and T.L. Steinmeier (2000). "Retirement in dual-career families: A structural model." Journal of Labor Economics **18**(3): 503-545.
44. Gustman, A.L. and T.L. Steinmeier (2004). "Social security, pensions and retirement behaviour within the family." Journal of Applied Econometrics **19**(6): 723-737.
45. Gustman, A.L. and T.L. Steinmeier (2005). "The social security early entitlement age in a structural model of retirement and wealth." Journal of Public Economics **89**(2-3): 441-463.

46. Gustman, A.L., T.L. Steinmeier and National Bureau of Economic Research. (2002). Retirement and the stock market bubble. Cambridge, Mass., National Bureau of Economic Research.
47. Gustman, A.L., T.L. Steinmeier and National Bureau of Economic Research. (2004a). Personal accounts and family retirement. Cambridge, MA, National Bureau of Economic Research.
48. Gustman, A.L., T.L. Steinmeier and National Bureau of Economic Research. (2004b). The Social Security retirement earnings test, retirement and benefit claiming. Cambridge, Mass., National Bureau of Economic Research.
49. Hammond, P. (1988). "Consequentalist foundations for expected utility." Theory and decision **25**: 25-78.
50. Haugen, F. (2000). Insentivvirkninger av skatte- og pensjonsregler. Oslo, Frischsenteret.
51. Heiss, F., M.D. Hurd, A. Börsch-Supan and National Bureau of Economic Research. (2003). Healthy, wealthy and knowing where to live : predicted trajectories of health, wealth and living arrangements among the oldest old. Cambridge, Mass., National Bureau of Economic Research.
52. Henretta, J., Angela M. O'Rand (1983). "Joint retirement in the dual worker family." Social forces **Volume 62:2 (December)**: 504-520.
53. Hernæs, E., Z. Jia and S. Strøm (2006). Retirement in non-cooperative and cooperative families. Oslo, Department of Economics, University of Oslo.
54. Hernæs, E., M. Sollie and S. Strøm (2000). "Early retirement and economic incentives." Scandinavian Journal of Economics **102**(3): 481-502.
55. Hernæs, E. and S. Strom (2000). "Empirical Specification of the model in 'Family Labour Supply When the Husband is Eligible for Early Retirement'." Frisch Center Working papers **1/2000**.
56. Hernæs, E. and S. Strom (2001). "Family labour supply when the husband is eligible for early retirement."
57. Heyma, A. (2004). "A structural dynamic analysis of retirement behaviour in the Netherlands." Journal of Applied Econometrics **19**(6): 739-759.
58. Hiedemann, B. (1998). "A Stackelberg model of social security acceptance decisions in dual career households." Journal of Economic Behavior & Organization **34**(2): 263-278.
59. Hurd, M. (1989). "The Joint Retirement Decision of Husbands and Wives." Social Security Bulletin **52**(1): 29-32.
60. Hurd, M.D. and K. McGarry (2002). "The predictive validity of subjective probabilities of survival." Economic Journal **112**(482): 966-985.
61. Iskhakov, F. (2003). Quasi-dynamic forward-looking model for joint household retirement decision under AFP scheme. Oslo, Department of Economics, University of Oslo.
62. Iskhakov, F. and M. Kalvarskaia (2003). AFP and OP data construction techniques. Oslo, Frischsenteret.
63. Jia, Z. (2000). Family labor supply when the husband is eligible for early retirement : some empirical evidences. Oslo, Department of Economics, University of Oslo.
64. Jia, Z. (2003). A mixture model of household retirement choice. Oslo, Department of Economics, University of Oslo.
65. Jia, Z. (2005). Retirement behavior of working couples in Norway : a dynamic programming approach. Oslo, Statistics Norway.
66. Kantorovich, L.V. (1976). "Mathematical-Methods in Economics." Problems of Economics **19**(4-6): 3-8.
67. 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.
68. Kreider, B. (1999). "Latent work disability and reporting bias." Journal of Human Resources **34**(4): 734-769.

69. Kreider, B. and R.T. Riphahn (2000). "Explaining applications to the US disability system - A semiparametric approach." Journal of Human Resources **35**(1): 82-115.
70. Krueger, A.B., B.D. Meyer and National Bureau of Economic Research. (2002). Labor supply effects of social insurance. Cambridge, MA., National Bureau of Economic Research.
71. Maddala, G.S. (1988). Limited dependent and qualitative variables in econometrics. Cambridge New York New Rochelle Melbourne Sydney, Cambridge University Press.
72. McGarry, K. (2002). Health and retirement : do changes in health affect retirement expectations? Cambridge, MA., National Bureau of Economic Research.
73. Michaud, P.C. and F. Vermeulen (2004). "A collective retirement model:Identification and estimation inthe presence of externalities." IZA Working paper series **1294**: 1-53.
74. Modernisert folketrygd. Bærekraftig pensjon for framtida: NOU 2004:1 (2004).
75. Murphy, K.M. and R.H. Topel (1985). "Estimation and Inference in 2-Step Econometric-Models." Journal of Business & Economic Statistics **3**(4): 370-379.
76. Olson, C.A. (1998). "A comparison of parametric and semiparametric estimates of the effect of spousal health insurance coverage on weekly hours worked by wives." Journal of Applied Econometrics **13**(5): 543-565.
77. Pension Reform - Safeguarding Our Pensions: summary of the report to Strorting nr. 12 (2004-2005) (2005).
78. Powers, E.T., D. Neumark and National Bureau of Economic Research. (2001). The supplemental security income program and incentives to take up Social Security early retirement : empirical evidence from matched SIPP and Social Security administrative files. Cambridge, MA., National Bureau of Economic Research.
79. Rhenius, D. (1974). "Incomplete Information in Markovian Decision Models." Annals of Statistics **2**(6): 1327-1334.
80. Røed, K. and F. Haugen (2003). "Early Retirement and Economic Incentives: Evidence from a Quasi-natural Experiment." Labour **17**(2): 203-228.
81. Røgeberg, O.J. (2000). "Married man and early retirement under AFP scheme." Memorandum of The Department of Economics, University of Oslo **02/2000**.
82. Rogowski, J. and L. Karoly (2000). "Health insurance and retirement behavior: evidence from the health and retirement survey." Journal of Health Economics **19**(4): 529-539.
83. Rust, J. (1987). "Optimal Replacement of Gmc Bus Engines - an Empirical-Model of Zurcher,Harold." Econometrica **55**(5): 999-1033.
84. Rust, J. (1990). Behavior of male workers at the end of the life cycle: an empirical analysis of states and control. Issues in the Economics of Aging. Chicago, University of Chicago Press**:** 317-383.
85. Rust, J. (1994). Structural estimation of Markov decision processes. Handbook of Econometrics, Vol. 4. D. L. M. R.F. Engle. Amsterdam, Elsevier Science B.V. **chap. 51:** 3081-3143.
86. Rust, J. (2006). Dynamic programming. New Palgrave Dictionary of Economics. S. D. a. L. Blume, Palgrave Macmillan.
87. Rust, J. and C. Phelan (1997). "How Social Security and Medicare affect retirement behavior in a world of incomplete markets." Econometrica **65**(4): 781-831.
88. Ruud, P.A. (1991). "Extensions of Estimation Methods Using the Em Algorithm." Journal of Econometrics **49**(3): 305-341.
89. Stock, J.H. and D.A. Wise (1990a). The pension inducement to retire: an option value analysis. **Chapter 7:** 205-229.
90. Stock, J.H. and D.A. Wise (1990b). "Pensions, the Option Value of Work, and Retirement." Econometrica **58**(5): 1151-1180.
91. Vermeulen, F. (2002). "Collective household models: Principles and main results." Journal of Economic Surveys **16**(4): 533-564.