AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Lassoing the Determinants of Retirement

Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen

CREATES Research Paper 2013-21

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Lassoing the Determinants of Retirement

Malene Kallestrup-Lamb[1,2], Anders Bredahl Kock[1,3], Johannes Tang Kristensen[1,4]

CREATES and Department of Economics and Business, Aarhus University,
Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark

This version: July 1, 2013.

Abstract

This paper uses Danish register data to explain the retirement decision of workers in 1990 and 1998. Many variables might be conjectured to influence this decision such as demographic, socio-economic, financially and health related variables as well as all the same factors for the spouse in case the individual is married. In total we have access to 399 individual specific variables that all could potentially impact the retirement decision. We use variants of the Lasso and the adaptive Lasso applied to logistic regression in order to uncover determinants of the retirement decision. To the best of our knowledge this is the first application of these estimators in microeconometrics to a problem of this type and scale. Furthermore, we investigate whether the factors influencing the retirement decision are stable over time, gender and marital status. It is found that this is the case for core variables such as age, income, wealth and general health. We also point out the most important differences between these groups and explain why these might be present.

Keywords: Retirement, Register data, High-dimensional data, Lasso, Adaptive Lasso, Oracle property, Logistic regression.
JEL classifications: C01, C25, J0, J14, J62.

Email addresses: mkallestup@econ.au.dk (Malene Kallestrup-Lamb), akock@creates.au.dk (Anders Bredahl Kock), jtang@creates.au.dk (Johannes Tang Kristensen)

[1]Support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation is gratefully acknowledged.

[2]Financial support from the The Danish Council for Independent Research, Social Sciences (11-105548/FSE) is gratefully acknowledged.

[3]Financial support from the The Danish Council for Independent Research, Social Sciences (11-116599/FSE) is gratefully acknowledged.

[4]Financial support from the The Danish Council for Independent Research, Social Sciences (11-116607/FSE) is gratefully acknowledged.

1. Introduction

Aging populations, progressive retirement behavior, increased flexibility with respect to retirement routes, and declining labor force participation among older workers over the last decades have increased the pressure on government expenses and is causing financial distress to the public pension system. This pattern is found in many advanced European countries such as France, the Netherlands, Italy, Germany, Britain and Sweden, see Blöndal and Scarpetta (1999); Ebbinghaus (2006); O'Rand and Henretta (1999). Thus, understanding which factors influence the retirement decision will be critical in understanding how the elderly workforce will evolve in the future and which policies to adopt to deal with the consequences of current retirement programs.

Building an econometric model for retirement requires important decisions regarding which variables to include. As competing economic theories might suggest different explanatory variables accommodating all of these might result in a vast set of potential explanatory variables. There is a wealth of evidence pointing to the relevance of a large host of demographic, socio-economic, financially and health related variables. See, e.g. Diamond and Hausman (1984); Antolin and Scarpetta (1998); Lindeboom (1998); Heyma (2004); Christensen and Kallestrup-Lamb (2012). For married individuals the relevance of the characteristics of the spouse adds to this complexity.

In this paper we consider a merged register-based data set consisting of a large, representative Danish sample of older workers drawn at random from the full population. We include a rich number of variables such as labor market status, level of education, age, occupation and sector variables, income, wealth, pension savings, and health. These are also available for the spouse for married individuals. Regarding the health variables, we have access to objective medical diagnosis codes for all patients who have been in contact with clinical hospital departments thereby avoiding the justification bias related to self-reported health measures, see Baker, Stabile, and Deri (2004) and Benítez-Silva, Buchinsky, Man Chan, Cheidvasser, and Rust (2004). Moreover, we consider information about hospital admissions, number of diagnoses, and number of treatments within a given year as well as visits to the general practitioner (GP).

Even though we have access to many variables that could be potentially explain the retirement decision only a subset of these might be relevant in explaining this decision. In large models traditional dimension reduction techniques such as testing or the application of information criteria can become computationally infeasible since the number of tests to be carried out and/or information criteria to be calculated increase exponentially in the number of variables. Furthermore, information criteria are known to be rather unstable, see e.g. Breiman (1996). These shortcomings have made shrinkage type estimators popular devices for selecting variables in high-dimensional models. The most prominent of these is the Least absolute shrinkage and selection operator of Tibshirani (1996). Since its inception many other estimators have been put forward. These include, but are not limited to, the smoothly clipped absolute deviation estimator of Fan and Li (2001), the Dantzig selector of Candes and Tao (2007), the bridge and marginal bridge estimators of Huang, Horowitz, and Ma (2008), and sure independence screening of Fan and Lv (2008). For recent overviews as well as many more references we refer the reader to Belloni and Chernozhukov (2011) and Bühlmann and van de Geer (2011).

The important feature in the shrinkage type estimators is that they perform estimation and variable selection at the same time. Furthermore, a lot of attention has been devoted to establishing the oracle property for most of the above procedures. This property entails establishing that all truly zero parameters are set exactly equal to zero with probability tending to one while this is not the case for any of the non-zero parameters. Furthermore, the non-zero parameters are estimated as efficiently as if only these variables had been included in the model from the outset—i.e. as if an oracle had revealed the true model prior to estimation. We shall elaborate further on this property in Section 4. In particular, we will use oracle efficient estimators to investigate which variables are important for the retirement decision. To the best of our knowledge this is the first application of these estimators in microeconometrics to a problem of this type and scale.

2. Institutional settings

In this section we give an overview of the institutional setting of the Danish labor market in order to understand the retirement options better. All exit routes out of the labour market are lumped into one retirement variable. We consider both old age pension as well as early retirement exit routes available to older workers. The latter include disability benefits, early retirement pay, civil service pension, and part-time pension. Unemployment insurance benefits and social assistance are not considered as separate routes of exit even though it is not uncommon for elderly workers to use unemployment as a retirement pathway, see Heyma (2004). Labor market pension schemes and private pension schemes are also described below but only considered a supplement to the early retirement schemes rather than an independent exit route. We proceed by giving further details on the individual exit routes.

Old-age pension is granted upon application from the age of 67 and conditional on at least 40 years of residence in Denmark between the ages of 18 and 67. It consists of a basic amount and a pension supplement and must be seen in conjunction with a number of other subsidies and benefits for which old-age pensioners may be eligible. These include favourable housing benefit rules for pensioners, support to heating expenses, a health allowance to the pensioner's own expenses for medicine, etc. Moreover, pensioners in general are entitled to a number of free services including hospital treatment, care in special residential accommodation, home care, physical maintenance training and rehabilitation. To these should be added a wide range of preventive and activating measures, such as cultural activities, teaching, physical exercise, etc.

Disability benefit is a tax-financed program assigned to individuals between the ages of 18 and 67 who are permanently unable to work and do not receive any other type of pension. Eligibility requires specific medical criteria to be met, assessed by a doctor, and conditioned on all possibilities to improve the applicant's labor market qualifications concerning rehabilitation, treatment, active social policy, etc. have been tried. The amount received depends on which type of disability pay is granted.

Early retirement pay is a voluntary labor market pension schemes that was introduced in 1979 as a labor market policy instrument. It offered workers between the ages of 60 and 66 the possibility to retire early and still maintain a reasonable income. It is not awarded on the basis of health conditions, but depends on the degree of labor market participation, type of membership of an approved unemployment insurance (UI) fund, and regular contributions for 10 to 25 years (depending on year of retirement). Thus, early retirement pay shares similarities with private pension schemes in a number of countries, including the U.S. Benefits are tied to previous wages, and employers also contribute to this retirement scheme. It is financially attractive, but unavailable once the disability route has been selected. Fully insured persons will receive 100% of the Unemployment Insurance benefit rate for the first two and a half years and afterwards a reduced 82%-rate for the rest of the period. By postponing the early retirement until the age of 63 (as of 1992) the member will receive the maximum rate for the whole period. Annuity payments from labor market pension schemes will induce a reduction in the early retirement pay by 60 percent, if paid out. For capital pensions no reduction is made.

Civil service pension is a statutory labor market pension scheme for civil servants financed through a pay-as-you-go system. This program is available from age the age of 60 to the age of

70. However other rules may apply if the civil servant was injured at work or suffered severe health problems. The size of the pension is based on the salary at the time of retirement and the length of the civil servant's employment period.

The part-time pension scheme gives people between 60 and 65 years of age not entitled to early retirement pay the possibility to reduce the number of hours worked per week. Different rules apply for this scheme depending on whether one is a wage earner or self employed mostly concerning previous and current connection to the labour market. A shift to a part-time job with the use of part-time pension could be a possible pathway to early retirement. However, even though many workers in Denmark express a desire to retire partly at the end of their working lives few people actually do so.

Labor market pension schemes and private pension schemes are considered a supplement to one of the retirement schemes described above. They can either be in the form of capital or an annuity. An annuity pension can either be discontinuous, ending after a pre-specified number of years with 10 being the minimum, or continuous thereby securing the individual a lifelong income stream independently of how long this person lives. Capital pension is paid as a lump sum from the age of 60 years at the earliest. The majority of labor market schemes are annuity based. In this paper we do not consider them as independent early retirement routes as most individuals have only made limited contributions in the sample period used for this study.

3. Data description

Next, we turn to describing the data set with particular emphasis on the type of explanatory variables available for the analysis of retirement decision. The full data base contains annual observations on all individuals in Denmark above 45 years of age for the period 1980 through 2001 with measurement in November each year. The data is based on administrative registers and contains no survey element. Hence, we reduce measurement errors, attenuation bias as well as justification bias, see e.g. Baker et al. (2004), Benítez-Silva et al. (2004) and Datta Gupta and Larsen (2010). We have information on various individual, demographic, financial, and socio-economic characteristics as well as health, and labor market status. We consider two different years, 1990 and 1998, which cover a period of very few reforms in the labour market regarding eligibilty for retirement. The sample used in this study consists of all individuals who are between 55 and 70 years old and active in the labor market. When analyzing 1990 we exclude individuals who are already retired as well as individuals that are unemployed for more than 47 weeks in a given year in any of the two years prior to 1990. The same rule is applied for 1998. Excluding the age group 45–54 avoids early retirement associated with limited job careers and loose labor market attachments.

Table 5 in the Appendix contains descriptive statistics of the explanatory variables for both 1990 and 1998. We investigate all individuals as well as married and singles separately. The category married includes both married and co-habiting individuals. In this section we comment on the descriptive statistics for married individuals in 1998 and remark if there are relevant differences in the other sub-samples. The total number of variables expected to affect the decision to retire amounts to 399 for married individuals. This includes the individuals' own characteristics as well as spouse characteristics.[5] Note that spouse variables are denoted with an (S). In order to maintain a more transparent structure the variables are grouped into 5 general categories: Personal Characteristics, Financial Indicators, Insurance & Pension, Employment, and Health. Finally, reference groups in the estimation are denoted by an (R).

For time-varying regressors we include variables for year $t-1$ (previous November) to explain retirement in year t, as we only observe that an exit to retirement has occurred sometime within a given year. This avoids a potential endogeneity issue arising if the value of a given characteristic is influenced by the retirement event. Note that gender, marital status, education, and region are considered time-invariant. Furthermore, we include variables for $t-2$ in order to take into account dynamic effects. However, when we consider variables for the spouse the levels are also included. Finally, most variables are normalized to the $[0,1]$ for estimation purposes.

We lump all exit routes out of the labour market into one retirement variable. The main routes are Disability, Early retirement pay, Civil service pension, Part-time pension and Old age pension. Due to the different nature of alternative exit routes we account for various eligibility specific explanatory variables in the estimation

Descriptive statistics for the dependent variable *retired* are shown in Table 1. We assess it across time as well as gender and marital status. Note that the number of retired individuals has decreased over time. However, there is a consistent gender specific pattern over time in that the percentage of retired females is higher. Furthermore, we see a higher proportion of singles classified as retired.

Next, we describe the explanatory variables in more detail. For the sake of readability we

[5]When assessing the sub-sample for singles or the sample for all individuals the spouse variables are not included.

Table 1. Detailed descriptive statistics for *retired*, gender-specific.

		Married		Single		All	
		Mean	SD	Mean	SD	Mean	SD
1990	All	0.091	0.288	0.115	0.319	0.096	0.295
	Male	0.084	0.277	0.097	0.296	0.086	0.280
	Female	0.104	0.306	0.130	0.336	0.112	0.315
1998	All	0.082	0.275	0.095	0.293	0.085	0.279
	Male	0.076	0.265	0.082	0.274	0.077	0.266
	Female	0.093	0.291	0.107	0.309	0.097	0.296

have gathered the explanatory variables in five broad categories. Each of them are described in turns in the sequel.

Personal Characteristics The first section of Table 5 covers Personal Characteristics. Three different groups of marital status are considered; *Married, Single* and *Co-habitation*. They take the value one if the individual is identified in one of the mutually exclusive groups and zero otherwise. In the subsequent analysis *married* and *co-habiting* are merged into one variable. In our sample, around 80% are married or cohabiting. *Male* is the gender dummy and we observe more males (63%) than females in the sample due the way the sample is constructed. The picture is very different for singles as only 46% of this subsample are males. The reasons are twofold. First of all, women on average have a lower income and single women are not hedged against a husbands income making it more difficult for them to retire early. Secondly, women have longer longevity and thus are overrepresented in the singles sample. The variable *Age* is restricted by our sampling criteria and ranges from 55–70 with a mean of 60 years. Furthermore, the age variable has been divided into five age groups: *Age 55–59*, *Age 60–61*, *Age 62–64*, *Age 65–66* and *Age 67–70*. This allow us to capture some age-specific effects related to different eligibility criteria in various retirement programs. Around 62% of the sample are in the age group 55–59 which is again explained by the conditioning on participation when the sample is selected. Married individuals dominate the younger ages whereas singles dominate the older age groups. Education is divided into five categories: Basic, Vocational, Short, Medium, and Long, and is defined by the individual's highest completed education level. Basic refers to primary or high school, only. Short, Medium, and Long are all higher educations beyond the high school level. Short and Medium refer to non-university degrees, with Short including less academic programs than Medium, and the latter typically requiring about 4 years after high school. Examples of educations under Short include real estate broker, actor, correspondent, technician with some training beyond vocational, laboratory worker, etc. Medium includes school teacher, journalist, librarian, accountant, nurse, midwife, social worker, army officer, some engineering, etc. Long includes all university degrees at the Bachelor level or higher, as well as engineers and architects with five years or longer programs. Since we look at cohorts of elderly persons, only 6% have a long education in our sample, and nearly 40% of the sample has only basic education, while about 38% has vocational training. Regarding geographical location we distinguish between eight different regions in Denmark. These are *Copenhagen*, *Greater Copenhagen*, *Zealand & Falster*,*Funen & Islands*, *Southern Jutland*, *Western Jutland*, *Central Jutland*, and *Northern Jutland*. Around 22% reside within the Copenhagen (the capital) metropolitan area and around 10–15% in each of the remaining areas.

For the spouse variables under Personal Characteristics we look at age and 8 different age categories. The age of the spouses range from 21 to 98 and thus the mean is slightly lower than for the individuals being analyzed. This also implies a broader grouping of the age variables. *Age <50*, *Age 50–54*, *Age 55–59*, *Age 60–61*, *Age 62–64*, *Age 65–66*, *Age 67–70*, and *Age >70*. As for the individual itself we again see that most individuals are in the category *Age 55–59* however for the spouse it only amounts to 36%. Furthermore, the age difference between married/co-habitating individuals is categorized into the following five groups: *Same age as spouse*, *Husband 1–4 years older*, *Husband more than 4 years older*, *Wife 1–4 years older*, and *Wife more than 4 years older*. In general, the husband is older – most often one to four years. The distribution of the education variables is similar to the one observed for the individual.

Financial Indicators The second section in Table 5 covers Financial Indicators. There are three main indicators; *Own income*, *Household income*, and *Wealth*. All these are deflated to year 2000 levels and measured in logarithms. *Own income* is the individuals income before taxes prior to any deductions for non-taxable income whereas *Household income* is measured after taxes. Due to the complex nature of the Danish tax-system it is sensible to include *Own income* as well as *Own income (S)* since their sum will still be different from *Household income*. Wealth is based on calculations from the tax authorities and is calculated as assets net of liabilities and hence includes the net value of real estate. Each of the three indicators is divided into five groups; *Low, Medium-Low, Medium, Medium-High*, and *High*. From the descriptive statistics we observe a general pattern for the three financial indicators. Between 3–7% fall into the low income group, 8% in the medium-low group, 14–19% in the medium group, and 21–27% in the medium-high group. The largest share is found in the highest groups. We see that 44% belong to the highest income group, 50% to the highest household income group and 32% to the highest wealth group. Moreover, we have an indicator variable for whether the individual is a home owner. Elderly home owners are increasingly becoming more reliant on their home equity as a source of retirement income. As 64% of the sample are home owners this variable could potentially be relevant. Finally, we see that *Own income* for the spouse is lower and the distribution between the groups is centered towards the lower groups compared to the working individuals in the sample.

Insurance & Pension Membership of a UI-fund is represented by the two variables *No unemployment insurance* and *Unemployment insurance*. It is part of the eligibility criteria for some of the retirement schemes, and membership exists for 77% of the sample. The *Supplementary labour market pension scheme,* represents payments paid out to the individual after the age of 67. All employees above the age 16 employed for more than nine hours a week pay contributions to to this scheme together with their employers. The wage earner pays one third of the full contribution. Only 2% of the sample receives these payments. Finally, we have information about contributions to private pension schemes. These types of schemes are considered a supplement to one of the retirement schemes and not an independent early retirement route as the main part of the individuals only have made a limited contribution over their working lives. The variables show how much the individual has contributed to the schemes in a given year making it a good indicator for how much the individual has put aside to supplement the public retirement schemes. Overall, we have two types of private pension schemes: *Private pension with annuity payments* and *Private pension with a capital payment.* These are each divided into three saving's categories: *None*, *Low*, and *High*. We see that around 30% of the sample is making contributions to a private annuity scheme and 45% is making contributions to a capital pension scheme. The numbers are slightly lower for singles.

Regarding the spouse variables it is seen that only 70% are members of a UI-fund and the number of spouses making contributions to a private pension scheme is around 10 percentage points lower. This is explained by the higher fraction of retired spouses.

Employment The extent of the labour market attachment is important due to different rules for full-time and part-time employed and whether or not they are insured. The indicator variables are divided into four groups: *Full-time employed & insured*, *Full-time employed & uninsured*, *Part-time employed & insured*, and *Part-time employed & uninsured*. These four variables classify 81% of the sample. The remaining 19% are captured by *Self employed* and *Assisting spouse* described below in the occupational specific indicators. From Table 5 we see that slightly more than 60% work full-time and are insured, whereas 8% of the full-time workers are uninsured. The split between insured and uninsured part-time workers is 5% in each category. We also note that the number of full-time workers is slightly higher for singles. The variables for experience are defined as the individual's work experience since 1980. Experience has been divided into five groups; *Job experience: <1 year*, *Job experience: 1–4 years*, *Job experience: 5–6 years*, *Job experience: 7–8 years*, and *Job experience: >8 years*. Note that the majority of workers (around 70%) have more than 8 years of experience in 1998. The yearly unemployment rate is based on the number of hours the individual has been unemployed relative to the number of possible hours worked. It may reflect multiple unemployment spells during the year and is divided into four groups; *Unemployed 1–3 months, Unemployed 3–6 months, Unemployed 6–9 months*, and *Unemployed 9–12 months*. The maximum number of weeks an individuals can be unemployed is 47 in order to still be considered as actively participating in the labour market. Around 5% of the sample is unemployed for 1–3 months during the year and the size of the remaining groups is less than 1.5% each. Finally, we note that singles have slightly higher unemployment rates.

Job characteristics are described through occupational indicators: *Self employed, Employed at high level, Employed at medium level, Employed at low level, Unskilled workers* and *Assisting spouse*. Finally, the dependent variable, *Retired,* is presented to illustrate that the sum of occupational indicators + retired sum to one. Among Employed workers, high level includes directors, managers, etc., medium level is other office personnel, and low level is skilled blue collar workers. These are broad categories, with 14% or more in each, except only 3% in *Assisting spouse*. The biggest group in the sample is classified as low level workers at around 34%. The sector specific variables are given by *Farming/Fishing, Manufacturing, Construction, Trade, Service, Hotel and Food, Transportation, Public* and *Unknown*. The last two variables represent the biggest part in this group.

The Employment section for the spouse reveals an interesting picture. Among the spouses there are less full-time workers and less uninsured, they are less experienced, and they experience spells of unemployment more often. Regarding occupational indicators we see that 31% of the spouses are retired in 1998. Furthermore, an extra group has been added called *Unemployed* classifying spouses that are unemployed more than 47 weeks a year. Almost 50% of the spouse have not been classified into one of the sector variables.

Health In addition to the standard background characteristics, we have information about the individual's health situation over time through several measures. These can be found in the Health section in Table 5. The indicator variable *Sickness benefits* takes the value one if the individual has received sickness pay during the year. This variable is intended to capture undiagnosed illnesses, thus complementing the indicators for diagnosis codes (see below). The first two days of illness are covered by the employer and around 7% of the sample experience longer sickness spells than two days and receive sickness benefits. The data for individual objective medical diagnosis codes is drawn from the Danish National Registry for Patients and includes information about admissions, actual diagnoses, treatments, and discharges for all patients who have been in contact with clinical hospital departments in Denmark during the sampling period. The essential feature of the data is that we have information about the **objective** medical diagnoses made at the time of a hospital discharge, and thereby avoid the justification bias related to self-reported health measures.

Within each year we have multiple observations for a given patient since the possibility of several admissions exists (approximately one third of the patients experience more than one admission within a given year). Furthermore, in relation to an admission, the patient is diagnosed with a main condition and possibly several additional conditions. The different diagnoses are organized according to WHO's international classification of diseases (ICD). From 1980 through 1993, ICD-8 is used, and from 1994 through 2001 ICD-10. This information is summarized in 14 dummy variables, each taking the value one if a person has been diagnosed with a disease in the associated category within the year. Both main and additional diagnoses are included, since it may be just as likely that it is an additional diagnosis that influences the decision to retire.

The categories we consider are: (1) *Malignant cancer* (includes leukemia, melanoma, and other malignant cancers); (2) *Benign tumors* (various types of tumors); (3) *Endocrine, nutritional, and metabolic diseases* (e.g., diabetes, obesity, etc.); (4) *Diseases of the blood and blood-forming organs* (nutritional and haemolytic anaemias);(5) *Mental and behavioral disorders* (dementia, delirium, schizophrenia, stress-related disorders, etc.); (6) *Diseases of the nervous system and sensory organs* (Alzheimer's, Parkinson's, epilepsy, sclerosis, migraine, apnoea, cataract, hearing loss, etc.); (7) *Diseases of the circulatory system* (ischaemic and other heart diseases, angina pectoris, acute rheumatic fever, high blood pressure, hypertension, stroke, etc.); (8) *Diseases of the respiratory system* (influenza, pneumonia, bronchitis, asthma, and other lung diseases); (9) *Diseases of the digestive system* (gastric ulcer, hernia, diseases of the liver and gallbladder, etc.); (10) *Diseases of the genitourinary system* (kidney stone, renal failure, other diseases of the urinary system and genital organs); (11) *Diseases of the skin* *and subcutaneous tissue* (infections of the skin, bullous disorders, urticaria and erythema); (12) *Diseases of the musculoskeletal system and connective tissue* (arthritis, osteoarthritis, Lyme disease, herniated disc, lumbago, osteoporosis, sclerosis, rheumatism, gout); (13) *Injury, poisoning, and other consequences of external causes* (bone fractures, dislocations, etc.); (14) *Other diseases*. The type of health event occurring most often is *Diseases of the circulatory system*, including stroke, at around 0.9%. This is followed by *Diseases of the digestive system*, including ulcer, at around 0.4%. In this respect it is important to stress that the individuals that we are observing are actively participating in the labour market. Therefore, they are less likely to suffer from a serious illness which would have forced them out of the labour market and therefore not be included in our sample. Moreover, we account for *Number of days of treatment*, *Number of diagnoses*, and *Number of admissions* within a given year. There are more admissions than either days of treatment or diagnoses within a given year, indicating that many admissions do not lead to any treatment, and that multiple admissions within a given year may lead to the same diagnosis.

In addition to the rich set of objective diagnosis codes, we have information about the number of services performed by the GP within a given year. More than one service can be carried out during a visit to the GP. The variable has been divided into four groups. *Doctor visits: 1–6 services*, *Doctor visits: 7–13 services*, *Doctor visits: 14–24 services*, and *Doctor visits: >24 services*. From Table 5 we see that 93% of the sample was in contact with the GP during the year and around 20% had more than 24 services performed.

Regarding the health indicators for the spouse we see a general pattern in terms of more health related problems. The type of health event occurring most often for the spouse is still *Diseases of the circulatory system*, but now the mean is almost three times as high at around 2.4%. This is followed by *Diseases of the digestive system* at 1.4%, *Diseases of the musculoskeletal system and connective tissue* at 1.2% and finally *Malignant cancer* at 1%. Moreover, we see that spouses are more likely to have longer treatment spells, higher number of diagnoses and admissions as well as doctor visits.

4. Methodology

In this section we give a short introduction to the penalized logistic regression to be used in modeling the retirement decision. The emphasis will be on the variable selection properties of these estimators. First, we introduce some notation. For a vector $x \in \mathbb{R}^p$ we shall let $\|x\| = \sqrt{\sum_{j=1}^{p} x_j^2}$ denote its ℓ_2-norm. For a set $A \subseteq \{1, ..., p\}$ the vector x_A denotes the subvector of x only consisting of the entries indexed by A. For a $p \times p$ matrix M, M_A denotes the submatrix only consisting of the rows and columns indexed by A. The symbol $\xrightarrow{p}$ shall signify convergence in probability while $\xrightarrow{d}$ denotes convergence in distribution.

4.1. Lasso-type estimators and the oracle property

Let Y be a random binary outcome variable with values in $\{0,1\}$. In the logistic regression the probability of an event $\{Y = 1\}$ occurring given a vector of explanatory variables X in $\mathbb{R}^p$ is modeled as

$$P(Y = 1|X = x) = F(x'\beta^*)$$

where $F(t) = (1 + e^{-t})^{-1}$ is the cumulative distribution function of the logistic distribution and β^* is a p-dimensional unknown parameter vector. This implies that for an independent sample of n observations the negative log-likelihood function is given by (see e.g. Heij, De Boer, Franses, Kloek, and Van Dijk (2004) for a text book treatment)

$$-\ell_n(\beta) = -\sum_{i=1}^{n} \left[y_i \log(F(x_i'\beta)) + (1 - y_i)\log(1 - F(x_i'\beta)) \right] \tag{1}$$

The parameter vector β^* may now be estimated by maximum likelihood[6]. This corresponds to minimizing (1). However, the minimizer $\hat{\beta}_{ML}$ will not posses any zeros while on the other hand it may be conjectured that only a (small) subset of the variables included in the model are truly relevant. In our study this corresponds to only a few of the many potential explanatory variable being relevant for explaining the retirement decision. Of course this lack of sparsity may be solved by standard techniques by testing whether a subset of the coefficients in β^* is zero by means of likelihood ratio (or similar) tests. But since each coefficient can be zero or not the number of sub models is as large as 2^p without any further prior knowledge on the parameter vector β^*. Furthermore, the final model one arrives at may depend on the order in which the sequence of tests is carried out. Similarly, if one wishes to use information criteria to select the correct model one has to estimate 2^p models which quickly becomes computationally infeasible for even moderate model sizes.

The above shortcomings of the standard likelihood based inference has lead to a great deal of research in estimators that perform estimation and variable selection simultaneously. The most common way of imposing sparsity on the model is by penalizing parameters that

[6]The maximum likelihood estimator exists and is unique under rather standard assumptions.

are different from zero. In particular, we shall focus on estimators that can be obtained as minimizers of objective functions of the form

$$L_n(\beta) = -\ell_n(\beta) + \lambda_n \sum_{j=1}^{p} w_j |\beta_j| \tag{2}$$

where λ_n is a positive sequence which determines the size of the penalty while $w_j,\ j = 1, ..., p$ are (potentially) data dependent weights. Note that (2) consists of two parts. The first part, $-\ell_n(\beta)$, is the negative log-likelihood function while the second part, $\lambda_n \sum_{j=1}^{p} w_j |\beta_j|$, penalizes parameters that are different from 0. The overall minimizer of $L_n(\beta)$ trades of these two parts and the tradeoff is determined by the size of λ_n. We will return to this issue later.

In recent years a lot of focus has been devoted to establishing the so-called *oracle property* of penalized estimators. Letting $\mathscr{A} = \{j : \beta_j^* \neq 0\}$ and $\mathscr{A}^c$ its complement, this entails showing that

Oracle Property:

1) $P(\hat{\beta}_{\mathscr{A}^c} = 0) \to 1$

2) $\sqrt{n}(\hat{\beta}_{\mathscr{A}} - \beta^*_{\mathscr{A}}) \tilde{\to} N\left(0, (I_{\mathscr{A}})^{-1}\right)$

where $I_{\mathscr{A}}$ denotes the Fisher information matrix for the relevant explanatory variables. 1) says that the estimators of all truly zero parameters will be set *exactly* equal to zero with probability tending to one. At this stage it is worth pointing out that this property is of course stronger than consistency of $\hat{\beta}$ which would imply $P\left(\|\hat{\beta}_{\mathscr{A}^c}\| > \epsilon\right) \to 0$ for all $\epsilon > 0$ but does not guarantee that any entry of $\hat{\beta}_{\mathscr{A}^c}$ will be set exactly equal to zero. Property 2) implies that $\hat{\beta}_{\mathscr{A}} \xrightarrow{p} \beta^*_{\mathscr{A}}$ which in turn means that no relevant variables will be excluded from the model asymptotically. In total, this implies that only relevant variables will be included in the model. Furthermore, 2) yields that the asymptotic distribution of the estimator of the non-zero coefficients is the same as if one had only included the relevant variables from the outset. Put differently, the non-zero coefficients are estimated as efficiently as if an oracle had revealed the true model prior to estimation and one had only included the relevant variables from the outset.

Let us next introduce the specific types of (2) that we consider in this paper and for each of these discuss if/when it possesses the oracle property.

If $w_j = 1$ for all $j = 1, ..., p$ in (2) one arrives at the Least Absolute Shrinkage and Selection Operator (Lasso) which was originally introduced by Tibshirani (1996) in the context of the linear regression model. Denote this minimizer by $\hat{\beta}_L$. The Lasso penalizes all parameters by an equal amount if they deviate from zero. In general it does not possess the oracle property and for this reason Zou (2006) developed the adaptive Lasso which chooses w_j more intelligently than the plain Lasso. In particular, the adaptive Lasso corresponds to $w_j = 1/|\tilde{\beta}_j|$ where $\tilde{\beta}_j$ is an initial estimator of the parameter β_j^*. Zou (2006) showed that for $\tilde{\beta} = \hat{\beta}_{ML}$ the adaptive Lasso $\hat{\beta}_{AL,ML}$ possesses the oracle property if $\lambda_n/\sqrt{n} \to 0$ and $\lambda_n \to \infty$ (as well as some further mild regularity conditions, see Zou (2006) Theorem 4)[7]. The condition $\lambda_n \to \infty$ is needed in order to penalize the truly zero parameters enough for the adaptive Lasso to shrink them to zero. On the other λ_n can't grow too fast either since this would imply non-zero parameters being set equal to zero. This is reflected in the requirement $\lambda_n/\sqrt{n} \to 0$.

An alternative route is to use the Lasso estimator $\hat{\beta}_L$ as initial estimator instead of $\hat{\beta}_{ML}$ in the adaptive Lasso. This corresponds to $\tilde{\beta} = \hat{\beta}_L$ and hence $w_j = 1/|\hat{\beta}_{L,j}|$ with the convention that $1/0 = \infty$ which of course implies that the $\hat{\beta}_{AL,L,j} = 0$ in case $\hat{\beta}_{L,j} = 0$. In practice, one simply leaves out the jth variable from the second step estimation when $\hat{\beta}_j = 0$ in the first step. Hence, using the Lasso as initial estimator implies that some variables are excluded in the first step as opposed to the case where $\hat{\beta}_{ML}$ is used as initial estimator. In practice this implies that $\hat{\beta}_{AL,L}$ is likely to be more sparse than $\hat{\beta}_{AL,ML}$. This can be useful in cases where one deals with many potential variables and wishes to reduce the dimension of the model a lot. Huang, Ma, and Zhang (2008) showed that under suitable regularity conditions $\hat{\beta}_{AL,L}$ possesses the oracle property.

Even though the two adaptive Lasso estimators above both possess the oracle property they can still suffer from finite sample biases. This motivates considering unpenalized estimation after model selection. Se Belloni and Chernozhukov (2013) for an example of this. In our case this corresponds to estimating β^* by maximum likelihood only including the non-zero entries of $\hat{\beta}_{AL,ML}$ or $\hat{\beta}_{AL,L}$, respectively. This also results in oracle efficient estimators since these estimators are asymptotically equivalent to the maximum likelihood estimator *only* including the relevant variables. This follows from the fact that $\hat{\beta}_{AL,ML}$ as well as $\hat{\beta}_{AL,L}$ will have the correct sparsity pattern asymptotically and so the third step maximum likelihood estimation is carried out only on the set of relevant variables. All results we report for the plain Lasso are also for post-selection estimated parameters. This still does not make the Lasso oracle efficient, however. The reason being that it does not select the correct sparsity pattern.

[7] Here $\hat{\beta}_{AL,ML}$ indicates that we are considering the Adaptive Lasso (AL) estimator with the Maximum Likelihood (ML) estimator used as initial estimator. A similar notation will be used in the sequel.

4.2. *Some caveats*

The oracle property is almost too good to be true. And in some sense it is. Hence, we also wish to point out some limitations to oracle efficient estimators. First and foremost, the above asymptotic results are all pointwise, i.e. derived for a fixed value of β^*. As argued in Leeb and Pötscher (2005) pointwise asymptotics may give a misleading picture of the finite sample distribution of the estimators. In particular, consistent model selection procedures will not be able to distinguish non-zero parameters from truly zero ones if the non-zero ones are sufficiently small. What we wish to convey with the above is that the oracle property should be interpreted and used with caution. For more details we refer to, e.g., Leeb and Pötscher (2008).

5. Implementation details

The results presented below have all been produced using `R` (R Core Team, 2012). Estimation of the standard logistic regression has been carried out using the built-in function `glm`. For the three different Lasso-based approaches the estimation is performed using the `glmnet` package (Friedman, Hastie, and Tibshirani, 2010). All variables are standardized internally in `glmnet` to ensure that any particular scaling of the data does not affect the results. The model is estimated for 100 values of λ_n chosen such that for the largest value no variables are included and for the smallest value most variables are included. The choice of which λ_n to use is then made by minimizing the Bayesian Information Criterion $\mathrm{BIC}_\lambda = -2\ell(\hat{\beta}_\lambda) + \mathrm{df}(\hat{\beta}_\lambda) \times \log n$ where $\mathrm{df}(\hat{\beta}_\lambda)$ is the number of non-zero coefficients in $\hat{\beta}$[8]. All models include intercepts, and in the case of the Lasso-based methods the intercept is not penalized.

The estimator of the covariance matrix is based on the standard Hessian

$$I_n = -\frac{\partial^2 \ell_n(\beta)}{\partial\beta\partial\beta'} = \sum_{i=1}^{n} F(x_i'\beta)(1-F(x_i'\beta))x_i x_i'$$

In the cases where post model selection estimation is carried out the post-estimation is performed again using the build-in function `glm` and hence standard errors are provided directly by `R` based on the above formula.

Even though the estimation problem is fairly straightforward the size of our data set does affect the computational burden considerably. To ease this burden we will therefore only use a subsample of the data for the estimation. This subsample is picked at random from

[8]Here $\hat{\beta}_\lambda$ is a the estimate of β^* pertaining to a particular λ. Note also that $\hat{\beta}$ can be *any* of the above Lasso-type estimators. Hence, we do not make any distinction in the notation at this point.

the entire data set. When possible we will used a sample size of 50,000 individuals. However, when considering only individuals who are single our data set does not provide enough observations to do so. Therefore in the following cases the sample size will differ from 50,000 and be: 1998/Single/Male: 31,706; 1998/Single/Female: 37,013; 1990/Single/Male: 30,133; 1990/Single/Female: 38,446. Clearly, the choice of this subsample will affect the estimation results, and the included number of variables in the Lasso methods will vary with both the given subsample and the size of the subsample. However, one would expect these variations to be concentrated around less important variables, whereas the truly important variables will always be included. This is similar to the way the choice of λ_n affects which variables are selected. As the penalty increases the model becomes smaller and less important variables are left out. To illustrate which variables are considered highly relevant the results below also show the effect of doubling the value of λ_n chosen by BIC on which variables are selected.

6. Results

Table 2 contains the results for a sub sample of 50,000 married individuals in 1998. Due to space considerations we only include variables in the table that are either significant at a 5% significance level in the logit model or deemed relevant by at least one of the shrinkage estimators. We again refer to the summary statistics in the Appendix for a list of all variables. It is sensible that the absolute value of the estimates increases when going from non-post estimation to post estimation since this step can reduce some of the finite sample shrinkage bias. Furthermore, it is worth noticing, that all procedures reduce the model size from 345 variables to between 20 and 40 variables. Hence, the dimension is reduced considerably.

We start by considering the personal characteristics. Note that, ceteris paribus, males are less likely to retire than females. This is a rather robust finding in the sense that even when λ_n (which determines the amount of shrinkage) is doubled the dummy for being a man is deemed relevant by all shrinkage estimators. The result that males work longer than females corresponds well with other findings in the literature, see e.g. Antolin and Scarpetta (1998) and Heyma (2004). It is also rather sensible that the likelihood to retire is increasing in the age of a person. All the included age variables are highly significant and remain in the model even when λ_n is doubled. In general the personal characteristics of the spouse do not play a significant role. However, the age of the spouse is included by Lasso-Post but without being significant. Moreover, we see that Lasso-Post is the only shrinkage method which includes an education variable, namely, *Education: Medium*. The fact that none of the other shrinkage methods include any education variables is surprising. However, the results in the literature regarding the effect of education on retirement is ambiguous. For example, Diamond and Hausman (1984) find that a higher level of education is associated with later retirement, whereas Lindeboom (1998) finds a positive effect of higher education on the retirement rate. Turning to the geographical variables, all shrinkage procedures find that the location of a household is immaterial to the retirement decision. This can be explained by the fact retirement laws and benefits are invariant across the country. Furthermore, Denmark is a rather small country which exhibits only minor regional differences. The full logit model, on the other hand, finds positive significant effects of living in Funen & Islands, South Jutland or North Jutland, which is in accordance with the results of An, Christensen, and Gupta (2004).

Table 2. Estimation results for married couples in 1998.

				AdaLasso(Logit)		AdaLasso(Lasso)	
	Variable	Full Logit	Lasso-Post	Non-post	Post	Non-post	Post
Personal Characteristics	Male	−0.368 (0.064)‡	**−0.349** (0.049)‡	**−0.233** (0.046)‡	**−0.347** (0.046)‡	**−0.352** (0.042)‡	**−0.368** (0.042)‡
	Age	0.056 (0.023)*	**0.267** (0.008)‡			**0.261** (0.007)‡	**0.263** (0.007)‡
	Age: 60–61	2.852 (0.098)‡	**2.075** (0.047)‡	**2.938** (0.056)‡	**3.058** (0.057)‡	**2.065** (0.047)‡	**2.079** (0.047)‡
	Age: 62–64	2.547 (0.148)‡	**1.245** (0.051)‡	**2.721** (0.061)‡	**2.880** (0.063)‡	**1.238** (0.051)‡	**1.251** (0.051)‡
	Age: 65–66	2.446 (0.210)‡	**0.566** (0.074)‡	**2.684** (0.082)‡	**2.895** (0.083)‡	**0.555** (0.074)‡	**0.572** (0.074)‡
	Age: 67–70	2.700 (0.265)‡		**2.984** (0.081)‡	**3.236** (0.083)‡		
	Education: Vocational	−0.117 (0.045)†					
	Education: Short	−0.290 (0.116)*					
	Education: Medium	0.217 (0.073)†	0.275 (0.057)‡				
	Region: Funen & Islands	0.154 (0.074)*					
	Region: South Jutland	0.221 (0.075)†					
	Region: North Jutland	0.186 (0.078)*					
	Age (S)	0.008 (0.014)	**−0.003** (0.004)				
Financial Indicators	Own income (L2)	−1.213 (0.935)		**−1.203** (0.510)*	**−1.324** (0.527)*		
	Own income: Medium-low (L1)	−0.455 (0.181)*					
	Own income: Medium (L1)	−0.627 (0.214)†					
	Own income: Medium-high (L1)	−0.934 (0.231)‡					
	Own income: High (L1)	−0.992 (0.247)‡	**−0.116** (0.072)	**−0.255** (0.071)‡	**−0.073** (0.073)		
	Own income: High (L2)	−0.449 (0.256)	**−0.323** (0.072)‡	−0.084 (0.072)	−0.291 (0.074)‡	**−0.384** (0.043)‡	**−0.386** (0.043)‡
	Household inc.: Medium-high (L1)	0.393 (0.166)*					
	Household inc.: High (L1)	0.417 (0.177)*					
	Wealth (L1)	1.004 (0.778)		**0.193** (0.098)*	**0.349** (0.100)‡		
	Wealth (L2)	−1.550 (0.708)*					
	Wealth: Medium (L2)	0.751 (0.368)*					
	Wealth: Medium-high (L2)	0.885 (0.416)*		0.011 (0.054)	0.041 (0.054)		
	Wealth: High (L2)	0.901 (0.453)*		**−0.027** (0.058)	**−0.030** (0.059)		
	Home owner (L1)	−0.225 (0.099)*					
	Home owner (L2)	0.219 (0.101)*					
	Own income (S)	1.989 (0.504)‡		**0.724** (0.224)†	**0.566** (0.226)*		
I & P	No unemp. insurance (L2)	−0.667 (0.196)‡	**−0.402** (0.055)‡	**−0.302** (0.053)‡	**−0.385** (0.055)‡	**−0.341** (0.054)‡	**−0.389** (0.055)‡
	Priv. pension, cap.: High (S)	−0.166 (0.079)*					

continued on the next page

Table 2. Estimation results for married couples in 1998 (continued).

				AdaLasso(Logit)		AdaLasso(Lasso)	
	Variable	Full Logit	Lasso-Post	Non-post	Post	Non-post	Post
Employment	Part-time emp., uninsured (L1)	0.571 (0.199)†	0.281 (0.073)‡	**0.215** (0.072)†	**0.293** (0.073)‡	**0.219** (0.073)†	**0.300** (0.073)‡
	Job experience: >8 years (L1)	−0.146 (0.536)	**0.152** (0.162)			**0.260** (0.158)	**0.182** (0.162)
	Job experience: >8 years (L2)	0.693 (0.520)	**0.287** (0.157)	**0.438** (0.053)‡	**0.456** (0.053)‡	**0.192** (0.153)	**0.298** (0.157)
	Unemployed: 9–12 months (L1)	0.867 (0.376)*		**0.504** (0.352)	**0.827** (0.332)*		
	Self employed (L1)	−0.181 (0.215)	**−0.366** (0.163)*			**−0.385** (0.162)*	**−0.366** (0.163)*
	Self employed (L2)	−0.297 (0.217)	**−0.415** (0.164)*	**−0.539** (0.073)‡	**−0.696** (0.077)‡	**−0.435** (0.164)†	**−0.420** (0.164)*
	Employed: Low level (L1)	0.342 (0.142)*	**0.235** (0.043)‡	**0.103** (0.041)*	**0.191** (0.041)‡	**0.168** (0.041)‡	**0.183** (0.041)‡
	Industry: Construction (L2)	−0.232 (0.105)*					
	Part-time emp., insured (S)	0.435 (0.199)*					
	Part-time emp., uninsured (S)	0.374 (0.162)*					
	Part-time emp., uninsured (L1)(S)	−0.418 (0.193)*					
	Unemployed: 6–9 months (S)	0.304 (0.152)*					
	Unemployed: 6–9 months (L2)(S)	−0.395 (0.148)†					
	Unemployed: 9–12 months (S)	0.482 (0.185)†					
	Retired (S)	3.850 (0.380)‡	**0.627** (0.040)‡	**1.204** (0.175)‡	**2.876** (0.344)‡	**0.614** (0.037)‡	**0.614** (0.038)‡
	Retired (L1)(S)	−0.481 (0.183)†		−0.089 (0.063)	−0.374 (0.062)‡		
	Self employed (S)	2.450 (0.402)‡		0.316 (0.180)	1.775 (0.347)‡		
	Unemployed (S)	3.406 (0.397)‡	0.594 (0.093)‡	**1.009** (0.188)‡	**2.541** (0.351)‡	**0.547** (0.094)‡	**0.582** (0.093)‡
	Unemployed (L1)(S)	−0.479 (0.207)*					
	Assisting spouse (L1)(S)	1.427 (0.386)‡					
	Industry: Trade (S)	−0.250 (0.094)†					
	Industry: Service (S)	−0.206 (0.099)*					
	Industry: Service (L2)(S)	0.224 (0.100)*					
	Industry: Unknown (S)	−2.959 (0.404)‡		**−0.510** (0.170)†	**−1.993** (0.342)‡		
	Industry: Unknown (L2)(S)	0.129 (0.052)*					
Health	Sickness benefits (L1)	1.227 (0.072)‡	**1.188** (0.059)‡	**1.121** (0.058)‡	**1.235** (0.059)‡	**1.187** (0.058)‡	**1.189** (0.059)‡
	Diag.: Benign tumors (L2)	−1.165 (0.722)		−0.036 (0.575)	−0.575 (0.626)		
	Diag.: Endocrine, etc. (L2)	1.164 (0.688)	0.883 (0.565)	**0.634** (0.581)	**1.231** (0.587)*	0.482 (0.573)	0.942 (0.562)
	Diag.: Blood (L1)	−1.846 (1.247)		**−1.139** (0.886)	**−2.020** (1.255)		
	Diag.: Blood (L2)	1.427 (2.391)		0.170 (1.777)	1.402 (1.763)		
	Diag.: Mental, behavioral (L2)	−10.37 (101.5)		**−3.201** (4.489)	**−10.97** (106.5)		
	Diag.: Circulatory system (L1)	0.373 (0.201)	0.375 (0.162)*			**0.349** (0.162)*	**0.382** (0.162)*
	Diag.: Respiratory system (L1)	−0.663 (0.426)		−0.165 (0.358)	−0.632 (0.415)		
	Diag.: Respiratory system (L2)	−1.313 (0.913)		**−0.497** (0.764)	**−1.367** (0.901)		
	Diag.: Digestive system (L2)	−0.852 (0.612)		−0.071 (0.497)	−0.820 (0.574)		
	# of days of treatment (L1)	7.417 (2.196)‡	**5.798** (1.422)‡	**7.748** (1.425)‡	**8.913** (1.485)‡	**6.074** (1.415)‡	**5.925** (1.419)‡
	# of days of treatment (L2)	8.584 (4.780)	**9.762** (2.575)‡	**11.29** (4.386)*	**9.556** (4.437)*	**10.71** (2.577)‡	**10.08** (2.584)‡
	# of admissions (L2)	1.871 (6.498)		0.367 (3.121)	3.459 (3.120)		
	Doctor visits: >24 services (L1)	0.188 (0.131)	0.141 (0.046)†			**0.136** (0.044)†	**0.174** (0.043)‡
	Doctor visits: >24 services (L2)	0.196 (0.107)	0.099 (0.048)*				
	Sickness benefits (L2)(S)	−0.272 (0.107)*					
	Diag.: Mental, behavioral (L1)(S)	0.813 (0.316)*		0.375 (0.313)	0.805 (0.293)†		
	Diag.: Skin (S)	−2.032 (1.030)*		**−1.130** (0.686)	**−2.069** (1.021)*		
	Doctor visits: 1–6 services (L1)(S)	0.588 (0.223)†		0.037 (0.044)	0.125 (0.044)†		
	Doctor visits: 7–13 services (L1)(S)	0.505 (0.225)*					
	Doctor visits: >24 services (L1)(S)	0.503 (0.228)*					

Notes: The results are based on a random sub-sample of the dataset consisting of 50,000 individuals. AdaLasso(Logit) refers to the Adaptive Lasso using the logit as initial estimator. Likewise, AdaLasso(Lasso) uses the Lasso as initial estimator. Some variables are included as lagged values from the previous two years, these are labelled (L1) or (L2). Variables pertaining to the spouse are labelled (S). Only variables that were selected by one of the Lasso methods or found to be significant in the full logit model are included in the table. The total number of variables is 345 of which Lasso, AdaLasso(Logit) and AdaLasso(Lasso) selected 25, 38 and 21, respectively. The tuning parameter λ_n is chosen using BIC. Bold indicates that the variable is still included when λ_n is doubled. The values in parenteses are standard errors and significance is indicated as: 5%(*), 1%(†), 0.1%(‡).

Regarding the financial indicators there can be two effects. First of all we have the substitution effect in that leisure is relatively more expensive for highly paid individuals indicating that they will retire later. On the other hand we have the income effect whereby more wealthy individuals save more and thus can afford to retire earlier. In Table 2 we see that for the income variables the substitution effect dominates, as the coefficients are negative. However, the income effect is controlled for by the inclusion of wealth which enters with a positive significant coefficient. Thus, a household which has accumulated much wealth over time does not have to stay in the labour market in order to accumulate sufficient wealth for retirement. Note that even though some of the categorical wealth variables are selected by the shrinkage methods they are not significant. In fact the majority of the wealth categories are deemed redundant and in particular it is of no importance whether one is a home owner or not. The latter is not surprising since it is the value of the house, not the fact that you own it that should matter, and this is included in the *wealth* variable. The full logit model, on the other hand, concludes that *home owner* is significant at both lags. However, the coefficients are of opposite signs and similar magnitude, hence this may merely be an artefact caused by lack of variation over time. The income of one's spouse has a positive and significant effect on the probability of retiring, thereby making it more affordable to retire early as a steady income stream is secured by the spouse. Note, none of the household income variables are deemed relevant. Recall that the household income is an after tax income while the *own income* variables are gross incomes. It might be surprising that the gross income variable is more relevant than the net income one. However, the former is more closely linked to the individual since it is the individuals *own*. Finally, it is interesting that most of these effects are only found for the Adaptive Lasso using the logit as initial estimator and not the other shrinkage procedures.

Turning to the Insurance & Pension category, all shrinkage procedures find that if one is without unemployment insurance then one is less likely to retire. The same is the case for the full logit model. This result seems reasonable as the attractive early retirement pay program requires membership of an UI-fund for a sufficiently long period of time. This is in accordance with the results of Christensen and Kallestrup-Lamb (2012). It is worth noticing that none of the many supplementary pension schemes that are included as explanatory variable are found to be relevant in predicting the retirement decision. Put differently, the models are rather sparse in this category.

Next, consider the group of employment variables. Greater labour market experience is associated with higher retirement probabilities. This seems reasonable as individuals who have participated over a longer period in the labour market have had time to build up retirement savings as well as contributions to pension funds. Furthermore, they are more likely to be eligible for the early retirement pay program as regular contributions to an UI-fund for 10–25 years is required. Thus, high experience gives individuals an extra opportunity for paid retirement. This finding remains even when the values of the tuning parameter λ_n are doubled.

The adaptive Lasso with the logit as initial estimator indicates that people who have been in the state of unemployment for 9–12 months in the previous year are more likely to retire. This can be explained by the fact that people with a loose connection to the labour market who are close to retirement age might choose to leave the labor force entirely instead of struggling with finding a new job for a short period of time. This is consistent with Lindeboom (1998). The same reasoning explains why all Lasso procedures find that people who are part time employed and uninsured are more likely to retire when compared to being full time employed and insured.

Regarding the occupational indicators in the employment group we see that compared to being employed at the high level being self-employed lowers the probability of retiring. This could be due to the fact that these people feel reluctant abandoning a company they have spent a large part of their lives building up. Note, however, that having a self employed spouse increases the probability of retiring. This can be explained by a labor sharing argument where the self employed spouse takes care of his or her company while the other part takes care of the household. Being a low level salaried worker compared to a high level one increases the probability of retirement. This is consistent with human capital theory and the empirical results of Heyma (2004). Having a retired spouse or an unemployed spouse both make one more likely to retire. This supports the theory of joint retirement where the former corresponds well with earlier empirical studies supporting the complementarities in leisure effect, see Henkens and Siegers (1991). Finally, it is worth noticing that it is not important which industry one (or one's spouse) works in.

In the health category, we see that receiving sickness benefits is important in explaining the retirement decision. This seems reasonable as it is a general indicator for poor health not captured by the objective diagnosis indicators below. Moreover, poor health increases the individual's uncertainty about their future in the labour market. This finding remains for all shrinkage procedures even after doubling the value of λ_n. We now consider the effects of health shocks as captured by the objective diagnosis indicators. Positive coefficients are expected under the assumption that health shocks may spur withdrawal from the labour market, and we do indeed find significant positive effects for *Endocrine, nutritional, and metabolic diseases* (e.g., diabetes, obesity, etc.), and *Diseases of the circulatory system* (ischaemic and other heart diseases, angina pectoris, acute rheumatic fever, high blood pressure, hypertension, stroke, etc.). These results are consistent with Christensen and Kallestrup-Lamb (2012). Note that especially the Lasso using the logit as initial estimator also includes further diagnosis indicators.

These are, however, not found to be significant. All procedures find that the longer treatments make retirement more likely. This seems reasonable as the length of treatment serves as a proxy for the severity of the illness. Moreover, we see that both lags of this variable are significant stressing the importance of the time dimension related to treatment and recovery. We realize that the individual's true health problems may not necessarily be captured by the objective diagnosis measures as certain conditions may be difficult to diagnose. Thus to account for this we include the number of services performed by the GP. This is found to be significant with an expected positive coefficient by Post-Lasso and the Lasso using the Lasso as initial estimator. Regarding the spouse variables we find an increased probability of entering retirement if the spouse is diagnosed with mental or behavioral disorders. Likewise, we find a positive effect for *Doctor visits: 1–6 services*. Surprisingly, however, we find a negative effect if the spouse is diagnosed with diseases of the skin and subcutaneous tissue.

6.1. Males, females, singles, and temporal robustness

So far we have considered married individuals in 1998. It is natural to ask whether the above findings are also valid for singles. Furthermore, it might be of interest to investigate whether the same variables determine the retirement decision for men and woman and if the relevant factors in 1998 are constant over time. To answer these questions we consider Table 3. This table contains the sign of the coefficients deemed non-zero by the adaptive Lasso with post estimation using the logit as initial estimator. For insignificant variables the sign is in a parenthesis. Note that we focus on this estimator here since the findings in Table 2 indicated that the variable selection is relatively robust across procedures with the adaptive Lasso using the logit as initial estimator selecting the most variables. We shall return to this robustness later.

In the personal characteristics category it is remarkable how stable the selected models over time, gender, and marital status are for the age variables as the same variables enter in the model across these dimensions. We find that in general the educational variables are more important for women than for men. Also there is a positive effect for single females in rural areas (Funen & Islands and North Jutland) in 1990.

Table 3. Estimation results for the Adaptive Lasso using the logit as initial estimator across samples..

		Married						Single						All					
		1990			1998			1990			1998			1990			1998		
		A	M	F	A	M	F	A	M	F	A	M	F	A	M	F	A	M	F
Personal Characteristics	Male				–												–		
	Age: 60–61	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Age: 62–64	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Age: 65–66	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Age: 67–70	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Education: Vocational															–			
	Education: Short									–						–	–		
	Education: Long									–						–			–
	Region: Zealand & Falster													+					
	Region: Funen & Islands									+				+		+			
	Region: North Jutland									+						+			
	Age: 65–66 (S)			–		(–)													
Financial Indicators	Own income (L1)	–	–	–		–	–			–	–	–		–	–	–		–	–
	Own income (L2)	–	(–)	–	–	–		–		–	–	–	–	–	–	(–)	–	–	
	Own income: Medium-low (L1)								–					–		–			
	Own income: Medium (L1)		+	–						–				–	(+)	–	(+)		
	Own income: Medium (L2)								–										
	Own income: Medium-high (L1)	–	–	–				–	–	–				–	–	–			
	Own income: Medium-high (L2)		+			+												+	–
	Own income: High (L1)	–	–	–	(–)	–		–	–	–		–		–	–	–	(–)	–	
	Own income: High (L2)				–		–				–		–				–		–
	Household income (L1)	+		+		+				+				+	+	+		+	
	Household income (L2)										+	+		(+)		(–)	+		
	Household inc.: Medium (L1)		+								+	+							
	Household inc.: Medium-high (L1)		+											+	+				
	Household inc.: Medium-high (L2)								+										
	Household inc.: High (L1)		+					+	+	+				+	+	+			
	Household inc.: High (L2)															–			
	Wealth (L1)				+	+	+			+	+		+				+		
	Wealth (L2)		(+)						(–)							+		(+)	
	Wealth: Medium-low (L2)		+			(–)			(+)			(+)		(+)	+	(+)			
	Wealth: Medium (L2)		(+)																
	Wealth: Medium-high (L2)		(+)		(+)	(+)									(+)			(+)	
	Wealth: High (L2)				(–)				(–)								(–)		
	Home owner (L1)							–	–	–	–		–	–	–	(–)	–		–
	Own income (S)	+	+	–	+	+	+												
	Own income: Medium-low (S)			+															
Insurance & Pension	No unemp. insurance (L1)					–													
	No unemp. insurance (L2)	–	–	–	–		–	–	–	–	–	–	–	–	–	–	–	–	–
	Supp. labour market pens. (L1)			–															
	Priv. pension, ann.: Low (L1)															–			
	Priv. pension, ann.: Low (L2)											–							
	Priv. pension, ann.: High (L2)											–							
	Priv. pension, cap.: High (L1)															–			
	Priv. pension, ann.: Low (S)					–													

continued on the next page

Table 3. Estimation results for the Adaptive Lasso using the logit as initial estimator across samples. (continued).

		Married						Single						All					
		1990			1998			1990			1998			1990			1998		
		A	M	F	A	M	F	A	M	F	A	M	F	A	M	F	A	M	F
Employment	Full-time emp., uninsured (L1)					+													
	Full-time emp., uninsured (L2)			−						−						−			
	Part-time emp., insured (L1)										+		(+)						(+)
	Part-time emp., insured (L2)					(+)												(+)	
	Part-time emp., uninsured (L1)		+		+	+		+	+	+	+	+		+	+	+	+	+	
	Part-time emp., uninsured (L2)		−												−				
	Job experience: 1–4 years (L1)									(+)									
	Job experience: 1–4 years (L2)		(+)						+	(+)				+	+	+			
	Job experience: 5–6 years (L1)		(+)					+		+				+	+				
	Job experience: 5–6 years (L2)	+	+	+		(+)			+	+				+	+	+			
	Job experience: 7–8 years (L1)	(+)	(+)											(+)	+				
	Job experience: 7–8 years (L2)	+	+	+		(+)	+	+	+	+				+	+	+	+	(+)	
	Job experience: >8 years (L1)							−		(−)									
	Job experience: >8 years (L2)				+	+	+				+	+	+				+	+	+
	Unemployed: 3–6 months (L1)					+								+		+			
	Unemployed: 6–9 months (L2)		+							+					+	+			
	Unemployed: 9–12 months (L1)		+		+		+	+	+		+	+	+		+		+	(+)	+
	Unemployed: 9–12 months (L2)	+	+	+		+								+	+	(+)		+	
	Self employed (L1)	−	−				−	−	−	−	−	−	−	−	−		−		−
	Self employed (L2)			−	−	−								(−)		−		−	
	Employed: Medium level (L2)													(−)					
	Employed: Low level (L1)				+					(−)						(−)	+		+
	Unskilled (L1)									+				+		+			
	Unskilled (L2)								+										
	Assisting spouse (L2)							+	(+)	+	+	(+)	+						
	Full-time emp., uninsured (S)		−																
	Part-time emp., uninsured (S)			+															
	Job experience: >8 years (L1)(S)						(+)												
	Unemployed: 6–9 months (L2)(S)					−													
	Unemployed: 9–12 months (L2)(S)					−													
	Retired (S)	+	+	+	+	+	+												
	Retired (L1)(S)			−	−														
	Retired (L2)(S)					−													
	Self employed (S)	+	+	−	+	+													
	Self employed (L1)(S)			+															
	Self employed (L2)(S)	+																	
	Unskilled (L2)(S)			+															
	Unemployed (S)	+	+		+	+													
	Assisting spouse (L1)(S)		+	(−)		+													
	Assisting spouse (L2)(S)						−												
	Industry: Unknown (S)	−	−		−	−													

continued on the next page

Table 3. Estimation results for the Adaptive Lasso using the logit as initial estimator across samples. (continued).

		Married						Single						All					
		1990			1998			1990			1998			1990			1998		
		A	M	F	A	M	F	A	M	F	A	M	F	A	M	F	A	M	F
Health	Sickness benefits (L1)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Diag.: Malignant cancer (L1)			−												−			
	Diag.: Malignant cancer (L2)										(−)	(−)	(−)						
	Diag.: Benign tumors (L1)			(−)									(−)						
	Diag.: Benign tumors (L2)				(−)	(−)												(−)	
	Diag.: Endocrine, etc. (L1)									+								+	
	Diag.: Endocrine, etc. (L2)			(+)	+	+				−	+		(+)				+	(+)	
	Diag.: Blood (L1)				(−)	−			(−)			(+)		(+)			(−)		(−)
	Diag.: Blood (L2)		(−)		(+)	+			(+)	(−)				(−)	(−)				(−)
	Diag.: Mental, behavioral (L1)			(+)				+	+	+	+	+				+			
	Diag.: Mental, behavioral (L2)	(−)	(−)	(−)	(−)	(−)	(−)	(+)		(+)		+	(−)	+	+	(+)			
	Diag.: Nervous system (L1)		(+)			+								+	+	(+)	+	+	
	Diag.: Nervous system (L2)		(+)	+						(+)		+				(+)			−
	Diag.: Circulatory system (L1)			+						(+)						(+)	+		
	Diag.: Circulatory system (L2)		(+)												(+)				+
	Diag.: Respiratory system (L1)				(−)														
	Diag.: Respiratory system (L2)			+	(−)	−	+				(+)					+			+
	Diag.: Digestive system (L1)										−	−							
	Diag.: Digestive system (L2)				(−)				(−)							(−)	−		−
	Diag.: Skin (L1)									(−)	+	(+)							
	Diag.: Skin (L2)						(−)		(−)								(−)	(−)	(−)
	Diag.: Musculoskeletal (L1)			+															
	Diag.: Musculoskeletal (L2)			(+)											(−)	+			
	Diag.: Injury, poisoning, etc. (L1)										−								
	Diag.: Injury, poisoning, etc. (L2)			(+)		(−)			+	(−)	+								
	Diag.: Other (L2)									(−)		−							
	# of days of treatment (L1)	+	+	+	+	+	+	+	+	+	+	+	(+)	+	+	+	(+)	+	+
	# of days of treatment (L2)			(−)	+	+	+				(−)	(−)				(−)	(+)	+	+
	# of diagnoses (L1)	(+)		(+)		(+)		(+)	(+)	−	+	(+)	+			(−)			+
	# of diagnoses (L2)			(+)		(−)	(−)		(−)	+	(−)	(−)	(−)	(+)	(+)	(−)	(+)	(−)	(−)
	# of admissions (L1)	(+)	(+)	(+)		(+)	(+)			(+)					(+)		(+)	(+)	
	# of admissions (L2)	(−)	−		(+)						+	+	(+)	(−)	(−)	(+)			
	Doctor visits: >24 services (L1)			+				+	+	+	+	+	+	+		+			+
	Doctor visits: >24 services (L2)																+		
	Diag.: Endocrine, etc. (L2)(S)		−																
	Diag.: Blood (L2)(S)					(−)													
	Diag.: Mental, behavioral (L1)(S)				+														
	Diag.: Nervous system (S)		−			−													
	Diag.: Genitourinary system (L1)(S)			−															
	Diag.: Skin (S)				−														
	Diag.: Skin (L1)(S)					(−)													
	# of days of treatment (L1)(S)					(+)													
	# of diagnoses (S)			−															
	# of diagnoses (L1)(S)		(+)				(+)												
	# of diagnoses (L2)(S)		(−)				(−)												
	# of admissions (S)	+	(+)	+															
	# of admissions (L1)(S)					(+)													
	Doctor visits: 1–6 services (L1)(S)				+														

Notes: The results are based on the different subsamples with the following abbrevations: All (A), Male (M), Female (F). A + or − sign indicates that variable was included and that the sign of the coefficient was positive or negative, respectively. Signs in parentheses indicate that the variables were not found to be significant at a 5% siginficance level in the post-estimated model.

Turning to the financial indicators we see that the finding from Table 2 that higher income decreases the likelihood of retiring is confirmed for both genders, for singles as well as in 1990. Even though household income was not selected by the shrinkage methods in Table 2 it is found to be relevant with a positive significant effect for a large number of the different samples. *Own income* of the spouse is, however, no longer selected for many of these samples.

Hence it could appear that the household income variables are capturing the effect we saw for spouse income in Table 2. For the wealth indicators it is of interest that in 1990 fewer wealth variables are relevant than in 1998. Consider for example the lagged wealth. It is relevant for both genders in 1998 while being irrelevant for both genders in 1990. Furthermore, we find that there is a negative effect on the probability of retirement for single home owners. This is in sharp contrast to married individuals for which home ownership is not found to be significant. The insurance and pension category is rather stable. Across all groups all signs are negative.

When considering the employment category we confirm the result that part time, uninsured individuals are more likely to retire across most samples. Furthermore, the positive effect of job experience is evident across the samples. However, lower categories are more important in 1990 due to the composition of the variables as being experience measured since 1980. Unemployment generally has a positive effect on retirement as also seen in Table 2. Across genders, time and marital status it is true that self employed individuals tend to postpone their retirement decision. In Table 2 we found that being a low level salaried worker compared a high level one increased the probability of retirement. This result is not found for singles. Moreover, we see a positive effect for unskilled individuals in 1990. Single females who have previously been an assisting spouse have an increased probability of retiring. In 1998 having an unemployed wife decreases the likelihood for men to retire. This can be because of a lower household income implying the need to stay in the labor market longer in order to save for retirement. The fact that having a retired spouse increases the probability of retiring is confirmed across gender and time. Having a self employed, unemployed or assisting spouse only increases the likelihood of retirement for male.

When considering the health category we notice that recipients of sickness benefits are more likely to retire irrespective of their gender, marital status and the year under consideration. Focussing on 1998 we find a positive effect on the probability of retirement for males from *Endocrine, nutritional, and metabolic diseases*, but a negative effect for singles from *Diseases of the digestive system*. In both years we find a positive effect for males from *Diseases of the nervous system and sensory organs*, a positive effect for married females for *Diseases of the circulatory system*, a positive effect for singles from both *Mental and behavioral disorders* and *Injury, poisoning, and other consequences of external causes*, and a positive effect for married females from *Diseases of the respiratory system*. For the latter effect we find the opposite for married males. Only in 1990 we find a negative effect from being diagnosed with malignant cancer for married females and a positive effect of diseases of the musculoskeletal system and connective tissue. The former result is consistent with the findings in Christensen and Kallestrup-Lamb (2012). We find very little evidence of effects from *Benign tumors*, *Diseases of the blood and blood-forming organs*, *Diseases of the skin and subcutaneous tissue*, and *Other diseases*. Turning to *# of days of treatment* this is seen to be a good indicator for early retirement across all categories. No general pattern is found for *# of diagnoses* and *# of admissions*. However, it is of interest that findings from Table 2 regarding the number of doctor visits being relevant for the retirement decision of married individuals seems to be driven entirely by women. Moreover, we see a consistent positive effect across years and gender for singles for *Doctor visits: >24 services*. No obvious pattern is identified for the health indicators for the spouse, except for a negative effect on the probability of retirement for males whose wives are diagnosed with diseases of the nervous system and sensory organs.

In Table 3 we only considered the post-estimated Adaptive Lasso using the logit as initial estimator. We will now gauge how robust the findings in Table 3 are across different models. Table 4 contains the fraction of overlap in the sign-pattern calculated as the number of entries in which the two vectors of coefficients have the same sign divided by the total length of the vector. In cases where the vectors being compared are of different dimension, only entries common to both vectors are being compared and the number of overlaps is divided by the potential number of overlaps.[9] All results in the table take the Adaptive Lasso using the logit as initial estimator as the reference. Hence the top part of the table illustates overlaps across all possible samples for this model only. Moving to the middle part of the table we instead compare the Adaptive Lasso using the logit as initial estimator to the Lasso for the various samples. Likewise, the bottom part of the table makes the comparison to the Adaptive Lasso using the Lasso as initial estimator.

Consider an example: Take the column Married; 1990; M; and the row Lasso; All; 1998; F; which has the value 0.58. Here we compare the overlap for the Adaptive Lasso using the logit as initial estimator for the sample of married males in 1990 to the Lasso for the sample of all females in 1998 and find that for 58% of the variables they agree on the sign of the coefficient and whether the variable should be included. Clearly, when comparing relatively different samples, such as this example, one would naturally expect to get a smaller overlaps. However, 0.58 is in fact the overall lowest value in the table indicating general stability of the findings. When looking across procedures, high overlaps are found for married individuals as

[9]The parameter vectors are of different length when we compare the estimated parameters for the subgroup of married individuals to the subgroup of singles as no spousal information is available for the latter.

Table 4. Sign-pattern match. Comparing the Adaptive Lasso using the logit as initial estimator to itself, the Lasso and the Adaptive Lasso using the Lasso as initial estimator..

				AdaLasso(Logit)																	
				Married						Single						All					
				1990			1998			1990			1998			1990			1998		
				A	M	F	A	M	F	A	M	F	A	M	F	A	M	F	A	M	F
AdaLasso(Logit)	Married	1990	A	**1.00**	**0.91**	0.90	**0.90**	0.88	**0.93**	0.89	0.80	0.77	0.80	0.82	0.86	0.84	0.88	0.73	0.81	0.86	0.81
			M	**0.91**	**1.00**	0.83	0.85	0.85	0.87	0.82	0.76	0.74	0.72	0.76	0.75	0.83	**0.93**	0.71	0.76	0.82	0.74
			F	0.90	0.83	**1.00**	0.84	0.84	0.87	0.81	0.74	0.77	0.78	0.76	0.80	0.76	0.77	0.78	0.76	0.81	0.75
		1998	A	**0.90**	0.85	0.84	**1.00**	0.88	**0.91**	0.82	0.78	0.68	0.81	0.78	0.86	0.73	0.76	0.67	0.87	0.85	0.82
			M	0.88	0.85	0.84	0.88	**1.00**	0.86	0.76	0.73	0.69	0.75	0.74	0.77	0.76	0.77	0.68	0.78	0.88	0.75
			F	**0.93**	0.87	0.87	**0.91**	0.86	**1.00**	0.86	0.80	0.73	0.86	0.82	0.90	0.76	0.80	0.68	0.88	0.88	0.89
	Single	1990	A	0.89	0.82	0.81	0.82	0.76	0.86	**1.00**	0.87	0.82	0.85	0.86	0.89	0.82	0.84	0.73	0.82	0.82	0.82
			M	0.80	0.76	0.74	0.78	0.73	0.80	0.87	**1.00**	0.73	0.78	0.78	0.80	0.78	0.78	0.72	0.80	0.76	0.80
			F	0.77	0.74	0.77	0.68	0.69	0.73	0.82	0.73	**1.00**	0.72	0.72	0.73	0.78	0.78	0.78	0.74	0.71	0.70
		1998	A	0.80	0.72	0.78	0.81	0.75	0.86	0.85	0.78	0.72	**1.00**	0.89	**0.92**	0.73	0.73	0.68	0.81	0.78	0.84
			M	0.82	0.76	0.76	0.78	0.74	0.82	0.86	0.78	0.72	0.89	**1.00**	0.85	0.76	0.75	0.70	0.77	0.78	0.79
			F	0.86	0.75	0.80	0.86	0.77	0.90	0.89	0.80	0.73	**0.92**	0.85	**1.00**	0.75	0.77	0.67	0.84	0.81	0.86
	All	1990	A	0.84	0.83	0.76	0.73	0.76	0.76	0.82	0.78	0.78	0.73	0.76	0.75	**1.00**	0.88	0.79	0.75	0.77	0.73
			M	0.88	**0.93**	0.77	0.76	0.77	0.80	0.84	0.78	0.78	0.73	0.75	0.77	0.88	**1.00**	0.74	0.79	0.82	0.75
			F	0.73	0.71	0.78	0.67	0.68	0.68	0.73	0.72	0.78	0.68	0.70	0.67	0.79	0.74	**1.00**	0.69	0.71	0.68
		1998	A	0.81	0.76	0.76	0.87	0.78	0.88	0.82	0.80	0.74	0.81	0.77	0.84	0.75	0.79	0.69	**1.00**	0.85	0.83
			M	0.86	0.82	0.81	0.85	0.88	0.88	0.82	0.76	0.71	0.78	0.78	0.81	0.77	0.82	0.71	0.85	**1.00**	0.80
			F	0.81	0.74	0.75	0.82	0.75	0.89	0.82	0.80	0.70	0.84	0.79	0.86	0.73	0.75	0.68	0.83	0.80	**1.00**
Lasso	Married	1990	A	**0.91**	0.85	0.87	0.88	0.85	**0.90**	0.85	0.78	0.72	0.78	0.78	0.83	0.79	0.78	0.72	0.83	0.82	0.80
			M	**0.90**	0.87	0.84	0.87	0.83	0.90	0.84	0.77	0.70	0.76	0.77	0.81	0.78	0.81	0.68	0.80	0.82	0.80
			F	0.89	0.82	0.89	0.87	0.82	0.89	0.80	0.72	0.70	0.76	0.76	0.79	0.72	0.72	0.72	0.78	0.77	0.78
		1998	A	**0.91**	0.85	0.86	**0.92**	0.85	**0.92**	0.85	0.77	0.72	0.81	0.79	0.85	0.75	0.74	0.68	0.86	0.83	0.83
			M	0.88	0.83	0.83	0.88	0.84	0.89	0.79	0.72	0.71	0.75	0.73	0.78	0.72	0.72	0.66	0.82	0.83	0.78
			F	**0.92**	0.85	0.86	**0.90**	0.84	**0.93**	0.86	0.77	0.70	0.80	0.78	0.86	0.75	0.74	0.66	0.81	0.80	0.84
	Single	1990	A	0.82	0.72	0.76	0.77	0.72	0.80	0.87	0.78	0.76	0.80	0.81	0.82	0.76	0.74	0.68	0.76	0.78	0.78
			M	0.86	0.76	0.78	0.82	0.76	0.84	0.87	0.84	0.72	0.80	0.79	0.83	0.75	0.78	0.68	0.83	0.81	0.80
			F	0.77	0.71	0.73	0.74	0.68	0.78	0.82	0.74	0.79	0.77	0.76	0.79	0.77	0.74	0.76	0.76	0.74	0.76
		1998	A	0.75	0.71	0.71	0.76	0.69	0.77	0.80	0.74	0.69	0.82	0.78	0.82	0.71	0.72	0.66	0.80	0.75	0.79
			M	0.86	0.77	0.80	0.80	0.75	0.85	0.88	0.78	0.75	0.82	0.83	0.84	0.78	0.78	0.70	0.80	0.80	0.81
			F	0.81	0.70	0.74	0.78	0.72	0.81	0.84	0.77	0.72	0.79	0.77	0.85	0.73	0.72	0.67	0.78	0.75	0.82
	All	1990	A	0.82	0.76	0.76	0.75	0.75	0.78	0.82	0.74	0.72	0.73	0.76	0.77	0.83	0.78	0.74	0.77	0.79	0.73
			M	0.78	0.79	0.70	0.69	0.72	0.74	0.76	0.72	0.68	0.68	0.70	0.70	0.79	0.82	0.68	0.74	0.77	0.70
			F	0.64	0.63	0.68	0.61	0.59	0.64	0.70	0.61	0.72	0.61	0.63	0.62	0.68	0.64	0.77	0.63	0.58	0.61
		1998	A	0.78	0.71	0.74	0.81	0.74	0.84	0.79	0.70	0.71	0.79	0.75	0.82	0.70	0.72	0.66	0.85	0.81	0.79
			M	0.83	0.74	0.77	0.82	0.78	0.84	0.82	0.75	0.72	0.77	0.75	0.81	0.75	0.74	0.68	0.81	0.86	0.81
			F	0.66	0.58	0.67	0.68	0.58	0.70	0.69	0.61	0.63	0.69	0.63	0.70	0.58	0.60	0.62	0.70	0.63	0.74
AdaLasso(Lasso)	Married	1990	A	**0.91**	0.85	0.87	0.88	0.85	**0.90**	0.86	0.78	0.72	0.78	0.78	0.84	0.80	0.78	0.73	0.82	0.83	0.80
			M	**0.92**	0.88	0.84	0.88	0.83	**0.92**	0.86	0.79	0.71	0.78	0.78	0.83	0.79	0.84	0.68	0.83	0.82	0.81
			F	0.90	0.83	**0.90**	0.88	0.83	0.90	0.82	0.74	0.69	0.77	0.78	0.80	0.73	0.73	0.71	0.78	0.80	0.79
		1998	A	**0.91**	0.85	0.87	**0.92**	0.85	**0.93**	0.86	0.78	0.72	0.82	0.80	0.87	0.75	0.75	0.68	0.86	0.84	0.85
			M	**0.92**	0.87	0.85	**0.90**	0.88	**0.92**	0.84	0.75	0.72	0.77	0.76	0.81	0.77	0.78	0.68	0.84	0.87	0.80
			F	**0.92**	0.85	0.86	**0.91**	0.84	**0.94**	0.87	0.79	0.71	0.82	0.80	0.88	0.76	0.76	0.66	0.83	0.81	0.86
	Single	1990	A	0.88	0.76	0.80	0.80	0.76	0.85	**0.92**	0.84	0.77	0.86	0.85	0.88	0.78	0.78	0.70	0.80	0.82	0.84
			M	0.88	0.76	0.80	0.84	0.78	0.86	**0.90**	0.84	0.73	0.83	0.82	0.86	0.78	0.78	0.70	0.82	0.83	0.83
			F	0.81	0.72	0.76	0.76	0.70	0.80	0.85	0.77	0.82	0.80	0.79	0.82	0.77	0.76	0.76	0.78	0.77	0.79
		1998	A	0.81	0.72	0.76	0.79	0.73	0.82	0.86	0.78	0.72	0.87	0.81	0.88	0.74	0.74	0.67	0.79	0.79	0.83
			M	0.89	0.80	0.82	0.81	0.79	0.88	**0.90**	0.81	0.76	0.82	0.84	0.85	0.78	0.81	0.70	0.82	0.82	0.82
			F	0.82	0.71	0.76	0.80	0.73	0.82	0.84	0.76	0.71	0.79	0.77	0.85	0.74	0.73	0.68	0.80	0.76	0.84
	All	1990	A	0.88	0.81	0.79	0.80	0.80	0.84	0.88	0.79	0.73	0.80	0.79	0.83	0.85	0.82	0.72	0.81	0.84	0.79
			M	0.83	0.82	0.73	0.74	0.73	0.82	0.84	0.77	0.70	0.75	0.77	0.78	0.80	0.85	0.69	0.80	0.79	0.75
			F	0.73	0.70	0.78	0.68	0.68	0.72	0.77	0.68	0.78	0.68	0.70	0.70	0.75	0.71	0.83	0.69	0.68	0.66
		1998	A	0.79	0.72	0.75	0.82	0.75	0.84	0.80	0.71	0.72	0.79	0.75	0.82	0.72	0.73	0.67	0.86	0.81	0.80
			M	0.86	0.76	0.78	0.85	0.79	0.86	0.84	0.76	0.71	0.78	0.76	0.82	0.75	0.77	0.66	0.84	0.87	0.81
			F	0.74	0.66	0.72	0.76	0.65	0.78	0.78	0.70	0.66	0.78	0.71	0.79	0.64	0.68	0.64	0.77	0.70	0.83

Notes: The following abbrevations are used: All (A), Male (M), Female (F). AdaLasso(Logit) refers to the Adaptive Lasso using the logit as initial estimator. Likewise, AdaLasso(Lasso) uses the Lasso as initial estimator. Values larger than 0.9 are in bold and values smaller than 0.7 are underlined.

the sparsity patterns quite often overlap by more than 90%. In particular, for females in 1998 we see that the Adaptive Lasso using the logit as initial estimator has a 94% overlap with the corresponding estimator using the Lasso to construct the weights. Turning to the models with the smallest overlap in the sparsity pattern one notices that these are very often found when considering females. However, this pattern is primarily found when we consider the sample of all females which could indicate the necessity to conduct separate analyses for married and single females.

7. Conclusions

Using a comprehensive Danish register data set we have investigated the factors driving the retirement decision of workers in 1990 and 1998. We find that age, several labour market indicators, income, wealth and a rich number of health variables are all very important. All the shrinkage procedures reduce the size of the model considerably. Furthermore, we investigate whether our findings are stable across gender and marital status. This is found to be the case for most variables. As another robustness check we experimented with doubling the value of the tuning parameter λ_n in order to investigate which variables are truly relevant. The variables found by the Lasso-type estimators are in accordance with earlier studies of the retirement decision. This shows that the use of shrinkage estimators give reasonable results and hence open the possibility to use these in future applied econometric research. Future avenues of research include extending the Lasso to settings of competing risk where the workers can retire into more than one state as well as more sophisticated ways of modelling the dynamics of the decision making process.

Appendix

Table 5. Descriptive statistics.

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Personal Characteristics	Married	0.957	0.204	0.000	0.000	0.757	0.429	0.943	0.232	0.000	0.000	0.753	0.431
	Co-habitation	0.043	0.204	0.000	0.000	0.034	0.182	0.057	0.232	0.000	0.000	0.046	0.209
	Single (R)	0.000	0.000	1.000	0.000	0.209	0.406	0.000	0.000	1.000	0.000	0.201	0.401
	Male	0.642	0.479	0.439	0.496	0.600	0.490	0.626	0.484	0.461	0.499	0.593	0.491
	Age	59.92	3.964	60.51	4.177	60.04	4.017	59.16	3.698	59.60	3.942	59.25	3.752
	Age: 55–59 (R)	0.536	0.499	0.475	0.499	0.523	0.499	0.617	0.486	0.572	0.495	0.608	0.488
	Age: 60–61	0.160	0.367	0.160	0.367	0.160	0.367	0.159	0.366	0.157	0.364	0.159	0.366
	Age: 62–64	0.153	0.360	0.169	0.375	0.156	0.363	0.120	0.325	0.137	0.343	0.123	0.329
	Age: 65–66	0.067	0.250	0.082	0.274	0.070	0.255	0.043	0.204	0.054	0.227	0.046	0.209
	Age: 67–70	0.084	0.278	0.114	0.318	0.090	0.287	0.060	0.238	0.079	0.270	0.064	0.245
	Education: Basic (R)	0.512	0.500	0.549	0.498	0.520	0.500	0.392	0.488	0.433	0.495	0.400	0.490
	Education: Vocational	0.325	0.468	0.277	0.448	0.315	0.464	0.381	0.486	0.326	0.469	0.370	0.483
	Education: Short	0.024	0.152	0.026	0.160	0.024	0.153	0.035	0.183	0.035	0.183	0.035	0.183
	Education: Medium	0.091	0.287	0.104	0.305	0.093	0.291	0.134	0.341	0.145	0.352	0.136	0.343
	Education: Long	0.049	0.216	0.043	0.203	0.048	0.213	0.059	0.236	0.062	0.241	0.060	0.237
	Region: Copenhagen (R)	0.215	0.411	0.329	0.470	0.239	0.427	0.183	0.386	0.304	0.460	0.207	0.405
	Region: Greater Copenhagen	0.117	0.321	0.109	0.312	0.115	0.319	0.146	0.353	0.134	0.340	0.143	0.350
	Region: Zealand & Falster	0.109	0.311	0.099	0.298	0.107	0.309	0.109	0.311	0.099	0.298	0.107	0.309
	Region: Funen & Islands	0.097	0.297	0.081	0.272	0.094	0.292	0.095	0.294	0.081	0.273	0.092	0.289
	Region: South Jutland	0.095	0.293	0.072	0.259	0.090	0.287	0.095	0.293	0.071	0.257	0.090	0.287
	Region: West Jutland	0.122	0.327	0.098	0.297	0.117	0.321	0.124	0.330	0.099	0.298	0.119	0.324
	Region: Central Jutland	0.153	0.360	0.134	0.340	0.149	0.356	0.159	0.366	0.140	0.347	0.155	0.362
	Region: North Jutland	0.092	0.289	0.077	0.267	0.089	0.285	0.089	0.285	0.073	0.259	0.086	0.280
	Age (S)	58.63	6.387					57.96	5.976				
	Age: <50 (S)(R)	0.066	0.248					0.057	0.232				
	Age: 50–54 (S)	0.187	0.390					0.211	0.408				
	Age: 55–59 (S)	0.314	0.464					0.362	0.481				
	Age: 60–61 (S)	0.118	0.323					0.121	0.326				
	Age: 62–64 (S)	0.141	0.348					0.121	0.327				
	Age: 65–66 (S)	0.067	0.249					0.049	0.216				
	Age: 67–70 (S)	0.074	0.262					0.054	0.225				
	Age: >70 (S)	0.033	0.177					0.025	0.156				
	Same age as spouse (S)(R)	0.080	0.272					0.093	0.290				
	Husband 1–4 years older (S)	0.425	0.494					0.462	0.499				
	Husband >4 years older (S)	0.347	0.476					0.291	0.454				
	Wife 1–4 years older (S)	0.116	0.320					0.122	0.327				
	Wife >4 years older (S)	0.031	0.173					0.032	0.176				
	Education: Basic (S)(R)	0.576	0.494					0.419	0.493				
	Education: Vocational (S)	0.293	0.455					0.373	0.484				
	Education: Short (S)	0.023	0.148					0.034	0.181				
	Education: Medium (S)	0.080	0.271					0.128	0.334				
	Education: Long (S)	0.028	0.166					0.045	0.208				
Financial Indicators	Own income (L1)	0.611	0.054	0.613	0.052	0.611	0.053	0.621	0.046	0.620	0.044	0.621	0.045
	Own income (L2)	0.611	0.055	0.612	0.057	0.611	0.056	0.621	0.043	0.621	0.040	0.621	0.043
	Own income: Low (L1)(R)	0.056	0.230	0.036	0.186	0.052	0.222	0.028	0.166	0.019	0.137	0.027	0.161
	Own income: Low (L2)(R)	0.060	0.237	0.045	0.206	0.057	0.231	0.028	0.164	0.019	0.136	0.026	0.159
	Own income: Medium-low (L1)	0.137	0.344	0.102	0.303	0.130	0.336	0.075	0.263	0.060	0.238	0.072	0.258
	Own income: Medium-low (L2)	0.135	0.341	0.104	0.305	0.128	0.334	0.074	0.262	0.060	0.238	0.071	0.257
	Own income: Medium (L1)	0.225	0.418	0.232	0.422	0.226	0.419	0.188	0.391	0.189	0.392	0.189	0.391
	Own income: Medium (L2)	0.210	0.407	0.209	0.406	0.210	0.407	0.192	0.394	0.194	0.395	0.193	0.394
	Own income: Medium-high (L1)	0.261	0.439	0.332	0.471	0.276	0.447	0.273	0.445	0.328	0.470	0.284	0.451
	Own income: Medium-high (L2)	0.254	0.435	0.329	0.470	0.270	0.444	0.272	0.445	0.330	0.470	0.284	0.451
	Own income: High (L1)	0.320	0.467	0.298	0.457	0.316	0.465	0.436	0.496	0.404	0.491	0.429	0.495
	Own income: High (L2)	0.341	0.474	0.314	0.464	0.336	0.472	0.434	0.496	0.398	0.489	0.427	0.495
	Household income (L1)	0.570	0.091	0.579	0.070	0.571	0.088	0.587	0.070	0.589	0.060	0.587	0.068
	Household income (L2)	0.568	0.095	0.577	0.076	0.570	0.092	0.591	0.070	0.594	0.054	0.592	0.067
	Household inc.: Low (L1)(R)	0.138	0.344	0.079	0.270	0.125	0.331	0.069	0.254	0.049	0.216	0.065	0.247
	Household inc.: Low (L2)(R)	0.145	0.352	0.092	0.289	0.134	0.341	0.063	0.243	0.044	0.206	0.059	0.236
	Household inc.: Medium-low (L1)	0.140	0.347	0.099	0.299	0.131	0.338	0.080	0.271	0.056	0.231	0.075	0.263
	Household inc.: Medium-low (L2)	0.138	0.345	0.102	0.302	0.131	0.337	0.065	0.246	0.045	0.206	0.061	0.239
	Household inc.: Medium (L1)	0.186	0.389	0.213	0.410	0.191	0.393	0.141	0.348	0.140	0.347	0.141	0.348

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Financial Indicators	Household inc.: Medium (L2)	0.175	0.380	0.200	0.400	0.180	0.385	0.098	0.297	0.094	0.291	0.097	0.296
	Household inc.: Medium-high (L1)	0.216	0.411	0.294	0.456	0.232	0.422	0.211	0.408	0.260	0.439	0.221	0.415
	Household inc.: Medium-high (L2)	0.212	0.409	0.289	0.453	0.228	0.419	0.174	0.379	0.203	0.402	0.180	0.384
	Household inc.: High (L1)	0.321	0.467	0.314	0.464	0.319	0.466	0.499	0.500	0.494	0.500	0.498	0.500
	Household inc.: High (L2)	0.329	0.470	0.318	0.466	0.327	0.469	0.600	0.490	0.615	0.487	0.603	0.489
	Wealth (L1)	0.492	0.253	0.521	0.236	0.498	0.250	0.483	0.259	0.486	0.260	0.484	0.259
	Wealth (L2)	0.484	0.257	0.511	0.242	0.490	0.254	0.479	0.263	0.482	0.260	0.479	0.262
	Wealth: Low (L1)(R)	0.213	0.410	0.172	0.377	0.204	0.403	0.233	0.423	0.232	0.422	0.233	0.423
	Wealth: Low (L2)(R)	0.225	0.417	0.186	0.389	0.217	0.412	0.241	0.428	0.234	0.424	0.240	0.427
	Wealth: Medium-low (L1)	0.103	0.304	0.090	0.286	0.100	0.300	0.084	0.277	0.080	0.271	0.083	0.276
	Wealth: Medium-low (L2)	0.104	0.306	0.089	0.285	0.101	0.302	0.085	0.279	0.087	0.282	0.086	0.280
	Wealth: Medium (L1)	0.139	0.346	0.142	0.349	0.140	0.347	0.144	0.351	0.142	0.349	0.144	0.351
	Wealth: Medium (L2)	0.137	0.344	0.142	0.349	0.138	0.345	0.138	0.345	0.143	0.351	0.139	0.346
	Wealth: Medium-high (L1)	0.199	0.399	0.230	0.421	0.206	0.404	0.220	0.414	0.222	0.416	0.220	0.414
	Wealth: Medium-high (L2)	0.198	0.398	0.227	0.419	0.204	0.403	0.218	0.413	0.225	0.418	0.219	0.414
	Wealth: High (L1)	0.345	0.475	0.367	0.482	0.350	0.477	0.320	0.466	0.324	0.468	0.320	0.467
	Wealth: High (L2)	0.336	0.472	0.355	0.479	0.340	0.474	0.317	0.465	0.310	0.463	0.316	0.465
	Home owner (L1)	0.576	0.494	0.572	0.495	0.575	0.494	0.640	0.480	0.608	0.488	0.633	0.482
	Home owner (L2)	0.575	0.494	0.565	0.496	0.573	0.495	0.618	0.486	0.588	0.492	0.612	0.487
	Own income (S)	0.566	0.124					0.595	0.086				
	Own income (L1)(S)	0.564	0.129					0.594	0.089				
	Own income (L2)(S)	0.563	0.134					0.595	0.090				
	Own income: Low (S)(R)	0.202	0.402					0.096	0.295				
	Own income: Low (L1)(S)(R)	0.202	0.401					0.094	0.292				
	Own income: Low (L2)(S)(R)	0.200	0.400					0.089	0.285				
	Own income: Medium-low (S)	0.196	0.397					0.174	0.379				
	Own income: Medium-low (L1)(S)	0.188	0.391					0.165	0.371				
	Own income: Medium-low (L2)(S)	0.183	0.387					0.157	0.364				
	Own income: Medium (S)	0.290	0.454					0.272	0.445				
	Own income: Medium (L1)(S)	0.288	0.453					0.277	0.447				
	Own income: Medium (L2)(S)	0.276	0.447					0.279	0.449				
	Own income: Medium-high (S)	0.157	0.364					0.198	0.398				
	Own income: Medium-high (L1)(S)	0.163	0.369					0.204	0.403				
	Own income: Medium-high (L2)(S)	0.167	0.373					0.209	0.407				
	Own income: High (S)	0.155	0.362					0.260	0.438				
	Own income: High (L1)(S)	0.159	0.366					0.261	0.439				
	Own income: High (L2)(S)	0.173	0.379					0.265	0.442				
Insurance & Pension	No unemp. insurance (L1)	0.318	0.466	0.308	0.462	0.316	0.465	0.233	0.423	0.264	0.441	0.239	0.427
	No unemp. insurance (L2)	0.327	0.469	0.316	0.465	0.325	0.468	0.226	0.418	0.255	0.436	0.232	0.422
	Unemp. insurance (L1)(R)	0.682	0.466	0.692	0.462	0.684	0.465	0.767	0.423	0.736	0.441	0.761	0.427
	Unemp. insurance (L2)(R)	0.673	0.469	0.684	0.465	0.675	0.468	0.774	0.418	0.745	0.436	0.768	0.422
	Supp. labour market pens. (L1)	0.031	0.173	0.079	0.269	0.041	0.198	0.023	0.151	0.033	0.179	0.025	0.157
	Supp. labour market pens. (L2)	0.020	0.140	0.054	0.225	0.027	0.162	0.016	0.124	0.024	0.153	0.017	0.131
	Priv. pension, ann.: None (L1)(R)	0.763	0.425	0.809	0.393	0.773	0.419	0.723	0.448	0.783	0.412	0.735	0.441
	Priv. pension, ann.: None (L2)(R)	0.756	0.430	0.804	0.397	0.766	0.423	0.719	0.449	0.782	0.413	0.732	0.443
	Priv. pension, ann.: Low (L1)	0.137	0.344	0.120	0.325	0.134	0.340	0.145	0.353	0.113	0.317	0.139	0.346
	Priv. pension, ann.: Low (L2)	0.139	0.345	0.123	0.328	0.135	0.342	0.150	0.357	0.118	0.322	0.144	0.351
	Priv. pension, ann.: High (L1)	0.098	0.298	0.070	0.255	0.092	0.289	0.131	0.337	0.103	0.304	0.125	0.331
	Priv. pension, ann.: High (L2)	0.104	0.305	0.072	0.259	0.097	0.296	0.129	0.335	0.100	0.300	0.123	0.329
	Priv. pension, cap.: None (L1)(R)	0.699	0.459	0.733	0.443	0.706	0.455	0.550	0.497	0.591	0.492	0.559	0.497
	Priv. pension, cap.: None (L2)(R)	0.738	0.440	0.765	0.424	0.743	0.437	0.554	0.497	0.592	0.491	0.562	0.496
	Priv. pension, cap.: Low (L1)	0.076	0.266	0.073	0.260	0.076	0.265	0.077	0.266	0.087	0.282	0.079	0.270
	Priv. pension, cap.: Low (L2)	0.081	0.273	0.077	0.267	0.080	0.272	0.083	0.277	0.095	0.294	0.086	0.280
	Priv. pension, cap.: High (L1)	0.220	0.414	0.191	0.393	0.214	0.410	0.369	0.482	0.318	0.466	0.358	0.480
	Priv. pension, cap.: High (L2)	0.176	0.381	0.153	0.360	0.172	0.377	0.358	0.479	0.308	0.462	0.348	0.476
	No unemp. insurance (S)	0.374	0.484					0.302	0.459				
	No unemp. insurance (L1)(S)	0.376	0.484					0.286	0.452				
	No unemp. insurance (L2)(S)	0.383	0.486					0.273	0.445				
	Unemp. insurance (S)(R)	0.626	0.484					0.698	0.459				
	Unemp. insurance (L1)(S)(R)	0.624	0.484					0.714	0.452				
	Unemp. insurance (L2)(S)(R)	0.617	0.486					0.727	0.445				

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Insurance & Pension	Supp. labour market pens. (S)	0.068	0.252					0.052	0.223				
	Supp. labour market pens. (L1)(S)	0.051	0.219					0.039	0.194				
	Supp. labour market pens. (L2)(S)	0.037	0.190					0.029	0.169				
	Priv. pension, ann.: None (S)(R)	0.803	0.398					0.768	0.422				
	Priv. pension, ann.: None (L1)(S)(R)	0.796	0.403					0.763	0.425				
	Priv. pension, ann.: None (L2)(S)(R)	0.801	0.399					0.761	0.426				
	Priv. pension, ann.: Low (S)	0.126	0.332					0.135	0.342				
	Priv. pension, ann.: Low (L1)(S)	0.132	0.339					0.138	0.345				
	Priv. pension, ann.: Low (L2)(S)	0.127	0.333					0.142	0.349				
	Priv. pension, ann.: High (S)	0.070	0.255					0.096	0.295				
	Priv. pension, ann.: High (L1)(S)	0.070	0.256					0.097	0.297				
	Priv. pension, ann.: High (L2)(S)	0.071	0.257					0.096	0.294				
	Priv. pension, cap.: None (S)(R)	0.747	0.435					0.650	0.477				
	Priv. pension, cap.: None (L1)(S)(R)	0.758	0.428					0.622	0.485				
	Priv. pension, cap.: None (L2)(S)(R)	0.788	0.408					0.620	0.485				
	Priv. pension, cap.: Low (S)	0.071	0.257					0.076	0.265				
	Priv. pension, cap.: Low (L1)(S)	0.073	0.260					0.081	0.273				
	Priv. pension, cap.: Low (L2)(S)	0.077	0.266					0.087	0.282				
	Priv. pension, cap.: High (S)	0.178	0.383					0.271	0.444				
	Priv. pension, cap.: High (L1)(S)	0.165	0.371					0.293	0.455				
	Priv. pension, cap.: High (L2)(S)	0.131	0.337					0.289	0.453				
Employment	Full-time emp., insured (L1)(R)	0.440	0.496	0.487	0.500	0.450	0.497	0.612	0.487	0.642	0.479	0.618	0.486
	Full-time emp., insured (L2)(R)	0.441	0.496	0.490	0.500	0.451	0.498	0.613	0.487	0.646	0.478	0.620	0.486
	Full-time emp., uninsured (L1)	0.135	0.342	0.142	0.349	0.136	0.343	0.084	0.278	0.091	0.288	0.086	0.280
	Full-time emp., uninsured (L2)	0.139	0.346	0.145	0.352	0.140	0.347	0.087	0.282	0.093	0.290	0.088	0.283
	Part-time emp., insured (L1)	0.087	0.282	0.069	0.253	0.083	0.276	0.050	0.218	0.029	0.168	0.046	0.209
	Part-time emp., insured (L2)	0.090	0.286	0.076	0.265	0.087	0.282	0.053	0.223	0.031	0.174	0.048	0.214
	Part-time emp., uninsured (L1)	0.087	0.282	0.111	0.314	0.092	0.289	0.060	0.237	0.076	0.265	0.063	0.243
	Part-time emp., uninsured (L2)	0.080	0.271	0.098	0.298	0.084	0.277	0.053	0.224	0.068	0.251	0.056	0.230
	Job experience: <1 year (L1)(R)	0.251	0.434	0.196	0.397	0.240	0.427	0.147	0.354	0.119	0.324	0.141	0.348
	Job experience: <1 year (L2)(R)	0.257	0.437	0.202	0.401	0.246	0.431	0.151	0.358	0.123	0.329	0.145	0.352
	Job experience: 1–4 years (L1)	0.118	0.323	0.119	0.324	0.118	0.323	0.064	0.244	0.072	0.258	0.065	0.247
	Job experience: 1–4 years (L2)	0.133	0.339	0.136	0.342	0.133	0.340	0.068	0.251	0.076	0.265	0.069	0.254
	Job experience: 5–6 years (L1)	0.107	0.309	0.112	0.315	0.108	0.311	0.036	0.187	0.039	0.194	0.037	0.189
	Job experience: 5–6 years (L2)	0.127	0.333	0.146	0.353	0.131	0.338	0.040	0.195	0.042	0.201	0.040	0.196
	Job experience: 7–8 years (L1)	0.118	0.323	0.152	0.359	0.125	0.331	0.041	0.199	0.044	0.205	0.042	0.200
	Job experience: 7–8 years (L2)	0.483	0.500	0.516	0.500	0.490	0.500	0.045	0.207	0.049	0.215	0.045	0.208
	Job experience: >8 years (L1)	0.405	0.491	0.421	0.494	0.409	0.492	0.712	0.453	0.726	0.446	0.715	0.451
	Job experience: >8 years (L2)							0.697	0.459	0.710	0.454	0.700	0.458
	Unemployed: 1–3 months (L1)	0.038	0.192	0.041	0.197	0.039	0.193	0.048	0.213	0.057	0.232	0.050	0.217
	Unemployed: 1–3 months (L2)	0.042	0.201	0.045	0.208	0.043	0.202	0.049	0.216	0.057	0.231	0.050	0.219
	Unemployed: 3–6 months (L1)	0.013	0.113	0.016	0.124	0.013	0.115	0.012	0.110	0.017	0.128	0.013	0.114
	Unemployed: 3–6 months (L2)	0.014	0.118	0.018	0.133	0.015	0.121	0.017	0.127	0.022	0.147	0.018	0.132
	Unemployed: 6–9 months (L1)	0.007	0.081	0.009	0.093	0.007	0.084	0.007	0.085	0.010	0.102	0.008	0.089
	Unemployed: 6–9 months (L2)	0.008	0.087	0.010	0.101	0.008	0.090	0.010	0.101	0.015	0.123	0.011	0.106
	Unemployed: 9–12 months (L1)	0.001	0.034	0.001	0.039	0.001	0.035	0.002	0.041	0.002	0.047	0.002	0.042
	Unemployed: 9–12 months (L2)	0.002	0.042	0.002	0.049	0.002	0.043	0.003	0.057	0.005	0.070	0.004	0.060
	Retired	0.091	0.288	0.115	0.319	0.096	0.295	0.082	0.275	0.095	0.293	0.085	0.279
	Self employed (L1)	0.197	0.398	0.182	0.386	0.194	0.395	0.162	0.369	0.156	0.363	0.161	0.368
	Self employed (L2)	0.197	0.397	0.181	0.385	0.193	0.395	0.163	0.370	0.155	0.362	0.162	0.368
	Employed: High level (L1)(R)	0.264	0.441	0.235	0.424	0.258	0.438	0.194	0.395	0.177	0.382	0.191	0.393
	Employed: High level (L2)(R)	0.265	0.441	0.236	0.424	0.259	0.438	0.193	0.395	0.176	0.381	0.189	0.392
	Employed: Medium level (L1)	0.166	0.372	0.247	0.431	0.183	0.387	0.131	0.337	0.141	0.348	0.133	0.339
	Employed: Medium level (L2)	0.167	0.373	0.249	0.432	0.184	0.387	0.133	0.339	0.146	0.353	0.135	0.342
	Employed: Low level (L1)	0.075	0.263	0.052	0.222	0.070	0.255	0.342	0.474	0.370	0.483	0.347	0.476
	Employed: Low level (L2)	0.075	0.263	0.051	0.221	0.070	0.255	0.344	0.475	0.369	0.483	0.349	0.477
	Unskilled (L1)	0.248	0.432	0.281	0.450	0.255	0.436	0.144	0.352	0.155	0.362	0.146	0.354
	Unskilled (L2)	0.247	0.431	0.279	0.448	0.253	0.435	0.140	0.347	0.152	0.359	0.142	0.349
	Assisting spouse (L1)	0.050	0.218	0.003	0.052	0.040	0.196	0.027	0.161	0.001	0.036	0.021	0.145
	Assisting spouse (L2)	0.050	0.219	0.004	0.066	0.041	0.198	0.027	0.162	0.002	0.046	0.022	0.146
	Industry: Farming/Fishing (L1)	0.020	0.139	0.018	0.133	0.019	0.138	0.031	0.173	0.032	0.176	0.031	0.174

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Industry: Farming/Fishing (L2)	0.020	0.139	0.019	0.137	0.019	0.138	0.031	0.174	0.031	0.173	0.031	0.174
	Industry: Manufacturing (L1)	0.095	0.293	0.089	0.285	0.094	0.292	0.070	0.255	0.067	0.251	0.069	0.254
	Industry: Manufacturing (L2)	0.094	0.292	0.092	0.289	0.094	0.292	0.070	0.255	0.068	0.252	0.070	0.254
	Industry: Construction (L1)	0.040	0.197	0.039	0.194	0.040	0.196	0.046	0.211	0.045	0.206	0.046	0.210
	Industry: Construction (L2)	0.040	0.197	0.039	0.193	0.040	0.196	0.046	0.210	0.046	0.209	0.046	0.210
	Industry: Trade (L1)	0.092	0.289	0.087	0.281	0.091	0.287	0.097	0.296	0.092	0.290	0.096	0.294
	Industry: Trade (L2)	0.092	0.288	0.087	0.282	0.091	0.287	0.097	0.296	0.091	0.288	0.096	0.294
	Industry: Service (L1)	0.084	0.277	0.080	0.272	0.083	0.276	0.099	0.298	0.094	0.292	0.098	0.297
	Industry: Service (L2)	0.083	0.276	0.082	0.274	0.083	0.276	0.098	0.298	0.095	0.293	0.098	0.297
	Industry: Hotel and Food (L1)	0.025	0.157	0.026	0.159	0.025	0.158	0.028	0.165	0.026	0.159	0.028	0.164
	Industry: Hotel and Food (L2)	0.025	0.157	0.026	0.160	0.025	0.158	0.028	0.165	0.027	0.163	0.028	0.164
	Industry: Transportation (L1)	0.049	0.217	0.046	0.210	0.049	0.215	0.041	0.198	0.037	0.189	0.040	0.197
	Industry: Transportation (L2)	0.049	0.217	0.046	0.208	0.049	0.215	0.041	0.197	0.039	0.193	0.040	0.197
	Industry: Public (L1)(R)	0.247	0.432	0.265	0.442	0.251	0.434	0.308	0.462	0.324	0.468	0.311	0.463
	Industry: Public (L2)(R)	0.250	0.433	0.257	0.437	0.251	0.434	0.310	0.462	0.319	0.466	0.311	0.463
	Industry: Unknown (L1)	0.347	0.476	0.349	0.477	0.348	0.476	0.283	0.450	0.286	0.452	0.283	0.451
	Industry: Unknown (L2)	0.346	0.476	0.352	0.478	0.348	0.476	0.282	0.450	0.288	0.453	0.283	0.451
	Full-time emp., insured (S)(R)	0.283	0.450					0.430	0.495				
	Full-time emp., insured (L1)(S)(R)	0.293	0.455					0.456	0.498				
	Full-time emp., insured (L2)(S)(R)	0.301	0.459					0.473	0.499				
	Full-time emp., uninsured (S)	0.061	0.240					0.047	0.212				
	Full-time emp., uninsured (L1)(S)	0.069	0.254					0.050	0.218				
	Full-time emp., uninsured (L2)(S)	0.075	0.263					0.053	0.224				
	Part-time emp., insured (S)	0.099	0.298					0.050	0.218				
	Part-time emp., insured (L1)(S)	0.108	0.311					0.056	0.231				
	Part-time emp., insured (L2)(S)	0.117	0.321					0.063	0.243				
	Part-time emp., uninsured (S)	0.066	0.248					0.036	0.188				
	Part-time emp., uninsured (L1)(S)	0.065	0.247					0.035	0.184				
Employment	Part-time emp., uninsured (L2)(S)	0.068	0.251					0.033	0.179				
	Job experience: <1 year (S)(R)	0.345	0.476					0.171	0.376				
	Job experience: <1 year (L1)(S)(R)	0.351	0.477					0.173	0.379				
	Job experience: <1 year (L2)(S)(R)	0.357	0.479					0.177	0.381				
	Job experience: 1–4 years (S)	0.187	0.390					0.097	0.296				
	Job experience: 1–4 years (L1)(S)	0.197	0.398					0.100	0.300				
	Job experience: 1–4 years (L2)(S)	0.212	0.408					0.104	0.306				
	Job experience: 5–6 years (S)	0.107	0.309					0.058	0.233				
	Job experience: 5–6 years (L1)(S)	0.137	0.344					0.060	0.238				
	Job experience: 5–6 years (L2)(S)	0.153	0.360					0.064	0.244				
	Job experience: 7–8 years (S)	0.128	0.335					0.066	0.248				
	Job experience: 7–8 years (L1)(S)	0.113	0.316					0.069	0.253				
	Job experience: 7–8 years (L2)(S)	0.278	0.448					0.072	0.258				
	Job experience: >8 years (S)	0.232	0.422					0.609	0.488				
	Job experience: >8 years (L1)(S)	0.202	0.401					0.598	0.490				
	Job experience: >8 years (L2)(S)							0.583	0.493				
	Unemployed: 1–3 months (S)	0.035	0.184					0.045	0.208				
	Unemployed: 1–3 months (L1)(S)	0.035	0.183					0.045	0.208				
	Unemployed: 1–3 months (L2)(S)	0.039	0.193					0.054	0.227				
	Unemployed: 3–6 months (S)	0.024	0.152					0.023	0.149				
	Unemployed: 3–6 months (L1)(S)	0.024	0.154					0.021	0.144				
	Unemployed: 3–6 months (L2)(S)	0.024	0.154					0.024	0.155				
	Unemployed: 6–9 months (S)	0.023	0.148					0.022	0.146				
	Unemployed: 6–9 months (L1)(S)	0.021	0.145					0.021	0.144				
	Unemployed: 6–9 months (L2)(S)	0.021	0.142					0.023	0.151				
	Unemployed: 9–12 months (S)	0.023	0.149					0.016	0.127				
	Unemployed: 9–12 months (L1)(S)	0.020	0.141					0.021	0.143				
	Unemployed: 9–12 months (L2)(S)	0.018	0.133					0.021	0.145				
	Retired (S)	0.329	0.470					0.312	0.463				
	Retired (L1)(S)	0.288	0.453					0.274	0.446				
	Retired (L2)(S)	0.256	0.437					0.246	0.431				
	Self employed (S)	0.103	0.304					0.091	0.288				
	Self employed (L1)(S)	0.109	0.312					0.094	0.292				

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Employment	Self employed (L2)(S)	0.113	0.316					0.097	0.296				
	Employed: High level (S)(R)	0.136	0.343					0.121	0.327				
	Employed: High level (L1)(S)(R)	0.145	0.352					0.125	0.331				
	Employed: High level (L2)(S)(R)	0.151	0.358					0.127	0.333				
	Employed: Medium level (S)	0.135	0.342					0.092	0.290				
	Employed: Medium level (L1)(S)	0.146	0.353					0.099	0.298				
	Employed: Medium level (L2)(S)	0.154	0.361					0.104	0.305				
	Employed: Low level (S)	0.026	0.159					0.225	0.418				
	Employed: Low level (L1)(S)	0.029	0.167					0.238	0.426				
	Employed: Low level (L2)(S)	0.031	0.173					0.249	0.433				
	Unskilled (S)	0.167	0.373					0.093	0.291				
	Unskilled (L1)(S)	0.181	0.385					0.096	0.295				
	Unskilled (L2)(S)	0.188	0.391					0.098	0.298				
	Unemployed (S)	0.046	0.210					0.037	0.189				
	Unemployed (L1)(S)	0.039	0.193					0.043	0.202				
	Unemployed (L2)(S)	0.039	0.193					0.044	0.205				
	Assisting spouse (L1)(S)	0.065	0.246					0.031	0.174				
	Assisting spouse (L2)(S)	0.068	0.251					0.033	0.179				
	Industry: Farming/Fishing (S)	0.017	0.129					0.015	0.121				
	Industry: Farming/Fishing (L1)(S)	0.010	0.097					0.017	0.129				
	Industry: Farming/Fishing (L2)(S)	0.010	0.098					0.017	0.128				
	Industry: Manufacturing (S)	0.052	0.222					0.042	0.201				
	Industry: Manufacturing (L1)(S)	0.046	0.209					0.038	0.191				
	Industry: Manufacturing (L2)(S)	0.046	0.209					0.038	0.191				
	Industry: Construction (S)	0.031	0.174					0.031	0.174				
	Industry: Construction (L1)(S)	0.022	0.148					0.024	0.154				
	Industry: Construction (L2)(S)	0.022	0.147					0.024	0.154				
	Industry: Trade (S)	0.070	0.255					0.075	0.263				
	Industry: Trade (L1)(S)	0.046	0.209					0.051	0.220				
	Industry: Trade (L2)(S)	0.046	0.209					0.051	0.219				
	Industry: Service (S)	0.070	0.255					0.067	0.249				
	Industry: Service (L1)(S)	0.044	0.206					0.052	0.221				
	Industry: Service (L2)(S)	0.044	0.205					0.051	0.221				
	Industry: Hotel and Food (S)	0.026	0.158					0.024	0.153				
	Industry: Hotel and Food (L1)(S)	0.013	0.114					0.015	0.123				
	Industry: Hotel and Food (L2)(S)	0.013	0.113					0.015	0.122				
	Industry: Transportation (S)	0.029	0.167					0.023	0.151				
	Industry: Transportation (L1)(S)	0.022	0.147					0.020	0.141				
	Industry: Transportation (L2)(S)	0.022	0.148					0.020	0.139				
	Industry: Public (S)(R)	0.166	0.372					0.255	0.436				
	Industry: Public (L1)(S)(R)	0.155	0.362					0.201	0.401				
	Industry: Public (L2)(S)(R)	0.153	0.360					0.200	0.400				
	Industry: Unknown (S)	0.540	0.498					0.469	0.499				
	Industry: Unknown (L1)(S)	0.642	0.479					0.583	0.493				
	Industry: Unknown (L2)(S)	0.645	0.479					0.585	0.493				
Health	Sickness benefits (L1)	0.080	0.271	0.084	0.278	0.081	0.272	0.072	0.258	0.071	0.258	0.072	0.258
	Sickness benefits (L2)	0.038	0.191	0.041	0.198	0.039	0.193	0.038	0.192	0.039	0.194	0.038	0.192
	Diag.: Malignant cancer (L1)	0.004	0.067	0.004	0.066	0.004	0.066	0.003	0.054	0.004	0.060	0.003	0.055
	Diag.: Malignant cancer (L2)	0.002	0.041	0.001	0.038	0.002	0.040	0.001	0.027	0.001	0.026	0.001	0.026
	Diag.: Benign tumors (L1)	0.003	0.058	0.003	0.056	0.003	0.058	0.002	0.045	0.002	0.044	0.002	0.045
	Diag.: Benign tumors (L2)	0.002	0.039	0.001	0.036	0.002	0.039	0.001	0.025	0.001	0.023	0.001	0.025
	Diag.: Endocrine, etc. (L1)	0.002	0.049	0.003	0.053	0.002	0.050	0.002	0.044	0.003	0.050	0.002	0.045
	Diag.: Endocrine, etc. (L2)	0.001	0.025	0.001	0.027	0.001	0.026	0.000	0.021	0.001	0.025	0.000	0.022
	Diag.: Blood (L1)	0.000	0.020	0.000	0.019	0.000	0.020	0.000	0.017	0.000	0.018	0.000	0.017
	Diag.: Blood (L2)	0.000	0.010	0.000	0.011	0.000	0.010	0.000	0.008	0.000	0.012	0.000	0.009
	Diag.: Mental, behavioral (L1)	0.001	0.023	0.001	0.030	0.001	0.024	0.000	0.016	0.001	0.035	0.000	0.021
	Diag.: Mental, behavioral (L2)	0.000	0.011	0.000	0.017	0.000	0.013	0.000	0.008	0.000	0.018	0.000	0.011
	Diag.: Nervous system (L1)	0.002	0.047	0.002	0.047	0.002	0.047	0.002	0.042	0.002	0.045	0.002	0.042
	Diag.: Nervous system (L2)	0.000	0.022	0.000	0.020	0.000	0.022	0.000	0.019	0.000	0.019	0.000	0.019
	Diag.: Circulatory system (L1)	0.009	0.095	0.008	0.092	0.009	0.094	0.009	0.093	0.008	0.090	0.009	0.092
	Diag.: Circulatory system (L2)	0.002	0.047	0.002	0.044	0.002	0.046	0.002	0.044	0.002	0.042	0.002	0.044

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Health	Diag.: Respiratory system (L1)	0.002	0.049	0.003	0.051	0.002	0.050	0.002	0.045	0.003	0.053	0.002	0.047
	Diag.: Respiratory system (L2)	0.001	0.025	0.001	0.024	0.001	0.025	0.001	0.023	0.001	0.024	0.001	0.023
	Diag.: Digestive system (L1)	0.006	0.075	0.005	0.072	0.006	0.074	0.004	0.065	0.005	0.070	0.004	0.066
	Diag.: Digestive system (L2)	0.001	0.034	0.001	0.037	0.001	0.035	0.001	0.029	0.001	0.031	0.001	0.030
	Diag.: Genitourinary system (L1)	0.004	0.064	0.004	0.064	0.004	0.064	0.003	0.052	0.003	0.057	0.003	0.053
	Diag.: Genitourinary system (L2)	0.001	0.031	0.001	0.032	0.001	0.031	0.001	0.022	0.001	0.028	0.001	0.024
	Diag.: Skin (L1)	0.001	0.025	0.001	0.024	0.001	0.025	0.000	0.022	0.001	0.023	0.000	0.022
	Diag.: Skin (L2)	0.000	0.012	0.000	0.011	0.000	0.011	0.000	0.011	0.000	0.011	0.000	0.011
	Diag.: Musculoskeletal (L1)	0.004	0.062	0.004	0.062	0.004	0.062	0.003	0.058	0.003	0.057	0.003	0.058
	Diag.: Musculoskeletal (L2)	0.001	0.029	0.001	0.026	0.001	0.028	0.001	0.026	0.001	0.030	0.001	0.027
	Diag.: Injury, poisoning, etc. (L1)	0.003	0.055	0.004	0.066	0.003	0.058	0.003	0.051	0.004	0.062	0.003	0.054
	Diag.: Injury, poisoning, etc. (L2)	0.001	0.025	0.001	0.029	0.001	0.026	0.000	0.022	0.001	0.027	0.001	0.023
	Diag.: Other (L1)	0.004	0.063	0.004	0.067	0.004	0.064	0.006	0.076	0.007	0.081	0.006	0.077
	Diag.: Other (L2)	0.001	0.030	0.001	0.035	0.001	0.031	0.001	0.036	0.001	0.039	0.001	0.037
	# of days of treatment (L1)	0.001	0.012	0.001	0.014	0.001	0.012	0.001	0.011	0.001	0.012	0.001	0.011
	# of days of treatment (L2)	0.000	0.006	0.000	0.007	0.000	0.006	0.000	0.005	0.000	0.006	0.000	0.005
	# of diagnoses (L1)	0.001	0.008	0.001	0.010	0.001	0.009	0.001	0.008	0.001	0.007	0.001	0.008
	# of diagnoses (L2)	0.000	0.005	0.000	0.006	0.000	0.005	0.000	0.005	0.000	0.004	0.000	0.004
	# of admissions (L1)	0.002	0.014	0.002	0.019	0.002	0.015	0.002	0.013	0.002	0.013	0.002	0.013
	# of admissions (L2)	0.001	0.008	0.001	0.012	0.001	0.009	0.000	0.008	0.000	0.007	0.000	0.007
	Doctor visits: 1–6 services (L1)	0.272	0.445	0.253	0.435	0.268	0.443	0.250	0.433	0.233	0.423	0.247	0.431
	Doctor visits: 1–6 services (L2)	0.239	0.427	0.218	0.413	0.235	0.424	0.240	0.427	0.212	0.409	0.234	0.424
	Doctor visits: 7–13 services (L1)	0.287	0.452	0.264	0.441	0.282	0.450	0.286	0.452	0.253	0.435	0.279	0.449
	Doctor visits: 7–13 services (L2)	0.288	0.453	0.265	0.441	0.283	0.450	0.281	0.450	0.248	0.432	0.275	0.446
	Doctor visits: 14–24 services (L1)	0.189	0.392	0.192	0.394	0.190	0.392	0.211	0.408	0.204	0.403	0.209	0.407
	Doctor visits: 14–24 services (L2)	0.195	0.396	0.195	0.397	0.195	0.396	0.200	0.400	0.194	0.395	0.198	0.399
	Doctor visits: >24 services (L1)	0.148	0.355	0.158	0.365	0.150	0.357	0.194	0.396	0.217	0.412	0.199	0.399
	Doctor visits: >24 services (L2)	0.116	0.320	0.126	0.332	0.118	0.323	0.177	0.382	0.201	0.401	0.182	0.386
	Sickness benefits (S)	0.080	0.271					0.073	0.260				
	Sickness benefits (L1)(S)	0.067	0.250					0.063	0.243				
	Sickness benefits (L2)(S)	0.050	0.218					0.046	0.209				
	Diag.: Malignant cancer (S)	0.011	0.106					0.010	0.098				
	Diag.: Malignant cancer (L1)(S)	0.013	0.112					0.010	0.101				
	Diag.: Malignant cancer (L2)(S)	0.013	0.115					0.011	0.102				
	Diag.: Benign tumors (S)	0.010	0.098					0.006	0.078				
	Diag.: Benign tumors (L1)(S)	0.009	0.095					0.006	0.076				
	Diag.: Benign tumors (L2)(S)	0.008	0.092					0.006	0.074				
	Diag.: Endocrine, etc. (S)	0.007	0.084					0.007	0.081				
	Diag.: Endocrine, etc. (L1)(S)	0.008	0.091					0.007	0.084				
	Diag.: Endocrine, etc. (L2)(S)	0.009	0.092					0.007	0.085				
	Diag.: Blood (S)	0.001	0.036					0.001	0.034				
	Diag.: Blood (L1)(S)	0.002	0.042					0.001	0.037				
	Diag.: Blood (L2)(S)	0.002	0.041					0.002	0.041				
	Diag.: Mental, behavioral (S)	0.002	0.046					0.002	0.039				
	Diag.: Mental, behavioral (L1)(S)	0.003	0.058					0.002	0.049				
	Diag.: Mental, behavioral (L2)(S)	0.004	0.065					0.003	0.056				
	Diag.: Nervous system (S)	0.007	0.082					0.006	0.080				
	Diag.: Nervous system (L1)(S)	0.008	0.090					0.006	0.080				
	Diag.: Nervous system (L2)(S)	0.009	0.093					0.006	0.079				
	Diag.: Circulatory system (S)	0.025	0.157					0.024	0.152				
	Diag.: Circulatory system (L1)(S)	0.027	0.161					0.024	0.152				
	Diag.: Circulatory system (L2)(S)	0.028	0.164					0.024	0.152				
	Diag.: Respiratory system (S)	0.009	0.096					0.008	0.088				
	Diag.: Respiratory system (L1)(S)	0.010	0.102					0.009	0.094				
	Diag.: Respiratory system (L2)(S)	0.011	0.106					0.010	0.098				
	Diag.: Digestive system (S)	0.016	0.127					0.014	0.117				
	Diag.: Digestive system (L1)(S)	0.017	0.130					0.014	0.119				
	Diag.: Digestive system (L2)(S)	0.017	0.130					0.015	0.120				
	Diag.: Genitourinary system (S)	0.017	0.130					0.011	0.105				
	Diag.: Genitourinary system (L1)(S)	0.016	0.124					0.011	0.102				
	Diag.: Genitourinary system (L2)(S)	0.015	0.122					0.010	0.101				

continued on the next page

Table 5. Descriptive statistics (continued).

		1990						1998					
		Married		Single		All		Married		Single		All	
	Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Health	Diag.: Skin (S)	0.002	0.045					0.002	0.041				
	Diag.: Skin (L1)(S)	0.002	0.046					0.002	0.043				
	Diag.: Skin (L2)(S)	0.002	0.048					0.002	0.044				
	Diag.: Musculoskeletal (S)	0.014	0.116					0.012	0.110				
	Diag.: Musculoskeletal (L1)(S)	0.013	0.114					0.012	0.107				
	Diag.: Musculoskeletal (L2)(S)	0.012	0.111					0.012	0.108				
	Diag.: Injury, poisoning, etc. (S)	0.011	0.102					0.010	0.101				
	Diag.: Injury, poisoning, etc. (L1)(S)	0.012	0.107					0.011	0.105				
	Diag.: Injury, poisoning, etc. (L2)(S)	0.012	0.111					0.012	0.108				
	Diag.: Other (S)	0.015	0.121					0.021	0.144				
	Diag.: Other (L1)(S)	0.017	0.128					0.023	0.149				
	Diag.: Other (L2)(S)	0.018	0.133					0.023	0.150				
	# of days of treatment (S)	0.004	0.022					0.003	0.018				
	# of days of treatment (L1)(S)	0.005	0.026					0.004	0.022				
	# of days of treatment (L2)(S)	0.005	0.028					0.004	0.023				
	# of diagnoses (S)	0.003	0.014					0.003	0.013				
	# of diagnoses (L1)(S)	0.003	0.015					0.003	0.013				
	# of diagnoses (L2)(S)	0.004	0.016					0.003	0.014				
	# of admissions (S)	0.006	0.022					0.005	0.022				
	# of admissions (L1)(S)	0.006	0.023					0.006	0.023				
	# of admissions (L2)(S)	0.006	0.025					0.006	0.024				
	Doctor visits: 1–6 services (S)	0.278	0.448					0.202	0.402				
	Doctor visits: 1–6 services (L1)(S)	0.232	0.422					0.215	0.411				
	Doctor visits: 1–6 services (L2)(S)	0.205	0.404					0.206	0.405				
	Doctor visits: 7–13 services (S)	0.262	0.440					0.269	0.444				
	Doctor visits: 7–13 services (L1)(S)	0.278	0.448					0.274	0.446				
	Doctor visits: 7–13 services (L2)(S)	0.279	0.448					0.270	0.444				
	Doctor visits: 14–24 services (S)	0.195	0.396					0.231	0.422				
	Doctor visits: 14–24 services (L1)(S)	0.212	0.409					0.224	0.417				
	Doctor visits: 14–24 services (L2)(S)	0.224	0.417					0.215	0.411				
	Doctor visits: >24 services (S)	0.197	0.397					0.259	0.438				
	Doctor visits: >24 services (L1)(S)	0.196	0.397					0.240	0.427				
	Doctor visits: >24 services (L2)(S)	0.165	0.371					0.227	0.419				

Notes: For the 1990 sample the “Married” subsample consists of 260,274 observations, the “Single” subsample of 68,579 observations, and the “All” subsample of 328,853 observations. For the 1998 sample the “Married” subsample consists of 273,141 observations, the “Single” subsample of 68,719 observations, and the “All” subsample of 341,860 observations. Some variables are included as lagged values from the previous two years, these are labelled (L1) or (L2). Variables pertaining to the spouse are labelled (S). For categorial dummies the reference group, i.e. the one omitted in the estimation, is labelled (R).

References

An, M. Y., B. J. Christensen, and N. D. Gupta (2004). Multivariate mixed proportional hazard modelling of the joint retirement of married couples. Journal of Applied Econometrics 19(6), 687–704.

Antolin, P. and S. Scarpetta (1998). Microeconomic analysis of the retirement decision: Germany. OECD Economics Department Working Papers 204, OECD Publishing.

Baker, M., M. Stabile, and C. Deri (2004). What do self-reported, objective, measures of health measure? Journal of Human Resources 39(4), 1067–1093.

Belloni, A. and V. Chernozhukov (2011). High dimensional sparse econometric models: An introduction. In P. Alquier, E. Gautier, and G. Stoltz (Eds.), Inverse Problems and High-Dimensional Estimation, Lecture Notes in Statistics, pp. 121–156. Springer Berlin Heidelberg.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. Bernoulli 19(2), 521–547.

Benítez-Silva, H., M. Buchinsky, H. Man Chan, S. Cheidvasser, and J. Rust (2004). How large is the bias in self-reported disability? Journal of Applied Econometrics 19(6), 649–670.

Blöndal, S. and S. Scarpetta (1999). The retirement decision in OECD countries. OECD Economics Department Working Papers 202, OECD Publishing.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. The Annals of Statistics 24(6), 2350–2383.

Bühlmann, P. and S. van de Geer (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer-Verlag, New York.

Candes, E. and T. Tao (2007). The Dantzig selector: Statistical estimation when p is much larger than n. The Annals of Statistics 35(6), 2313–2351.

Christensen, B. J. and M. Kallestrup-Lamb (2012). The impact of health changes on labor supply: Evidence from merged data on individual objective medical diagnosis codes and early retirement behavior. Health Economics 21(S1), 56–100.

Datta Gupta, N. and M. Larsen (2010). The impact of health on individual retirement plans: Self-reported versus diagnostic measures. Health Economics 19(7), 792–813.

Diamond, P. and J. Hausman (1984). The retirement and unemployment behavior of older men. In H. Aaron and G. Burtless (Eds.), Retirement and Economic Behavior. Washington DC: The Brookings Institution.

Ebbinghaus, B. (2006). Reforming Early Retirement in Europe, Japan and the USA. Oxford University Press.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96(456), 1348–1360.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(5), 849–911.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33(1), 1–22.

Heij, C., P. De Boer, P. H. Franses, T. Kloek, and H. K. Van Dijk (2004). Econometric methods with applications in business and economics. Oxford University Press, Oxford.

Henkens, K. and J. Siegers (1991). The decision to retire: The case of dutch men aged 50–64. European Journal of Population/Revue européenne de Démographie 7(3), 231–249.

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

Huang, J., J. L. Horowitz, and S. Ma (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. The Annals of Statistics 36(2), 587–613.

Huang, J., S. Ma, and C.-H. Zhang (2008). The iterated lasso for high–dimensional logistic regression. Technical Report 392, University of Iowa, Department of Statistics and Actuarial Science.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. Econometric Theory 21(1), 21–59.

Leeb, H. and B. M. Pötscher (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. Journal of Econometrics 142(1), 201–211.

Lindeboom, M. (1998). Microeconomic analysis of the retirement decision: the Netherlands. OECD Economics Department Working Papers 207, OECD Publishing.

O'Rand, A. M. and J. C. Henretta (1999). Aging in the welfare state: strategic crossnational comparisons of life course variability and inequality. In Age and Inequality: Diverse Pathways Through Later Life, pp. 186–206. Westview Press Boulder, CO.

R Core Team (2012). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58(1), 267–288.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101(476), 1418–1429.

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2013-04: Tom Engsted and Thomas Q. Pedersen: Housing market volatility in the OECD area: Evidence from VAR based return decompositions

2013-05: Søren Johansen and Bent Nielsen: Asymptotic analysis of the Forward Search

2013-06: Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets

2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns

2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation

2013-09: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox

2013-10: Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models

2013-11: Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis

2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications

2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions

2013-14: Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments

2013-15: Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation

2013-16: Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression

2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model

2013-18: Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models

2013-19: Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange

2013-20: Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models

2013-21: Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement