

18-09

Understanding Joint Retirement

CAHIER DE RECHERCHE
WORKING PAPER

Pierre-Carl Michaud, Arthur Van Soest and Luc Bissonnette

Septembre / September 2018



UNIVERSITÉ
LAVAL

Faculté des sciences sociales

HEC MONTRÉAL

ESG UQÀM



La Chaire de recherche Industrielle Alliance sur les enjeux économiques des changements démographiques est une chaire multi-institutionnelle qui s'appuie sur un partenariat avec les organisations suivantes :

- **Centre interuniversitaire de recherche en analyse des organisations (CIRANO)**
- **iA Groupe financier**
- **Retraite Québec**

Les opinions et analyses contenues dans les cahiers de recherche de la Chaire ne peuvent en aucun cas être attribuées aux partenaires ni à la Chaire elle-même et elles n'engagent que leurs auteurs.

Opinions and analyses contained in the Chair's working papers cannot be attributed to the Chair or its partners and are the sole responsibility of the authors.

© 2018 Pierre-Carl Michaud, Arthur Van Soest and Luc Bissonnette. Tous droits réservés. All rights reserved. Reproduction partielle permise avec citation du document source, incluant la notice ©. Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.

Dépôt légal : Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada, 2018.
ISSN 2368-7207



Understanding Joint Retirement^a

Pierre-Carl Michaud^b Arthur Van Soest^c Luc Bissonnette^d

September 1, 2018

Abstract

Evidence from different sources shows that spouses' retirement decisions are correlated. Retirement policies affecting individuals in couples are therefore also likely to affect behavior of their spouses. It is therefore important to account for joint features in modeling retirement. This paper studies a structural collective model of labor supply and retirement of both partners in a couple with interdependent preferences, imperfect knowledge of preferences of the spouse, and subjective expectations about the future. We propose a novel method to estimate preferences and the intra-household bargaining process, which relies on stated preferences data collected in the Health and Retirement Study. Respondents were asked to choose between hypothetical retirement trajectories describing the retirement ages and replacement rates of both spouses from three perspectives: considering their own preferences only, the preferences of their spouse only, or the most likely decision for the household. With these data, all model parameters are identified and potential sources of joint retirement can be disentangled. We find that males misperceive their wives' preferences, overestimating their disutility of work. Our estimates correct for this bias. They suggest that correlation in unobserved heterogeneity components of the partners' marginal utility of leisure explains a large share of joint retirement decisions. We also find significant positive complementarities in leisure, but this explains a much smaller part of joint retirement.

Keywords: Collective models, Leisure complementarity, Stated choices, Identification

JEL Codes: D13, J26, C81

^aWe acknowledge financial support from the National Institute of Aging (NIA) and the Netherlands Scientific Organization of Scientific Research (NWO).

^bHEC Montreal and NBER. Email: pierre-carl.michaud@hec.ca

^cNetspar, Tilburg University, RAND and IZA. Email: A.H.O.vanSoest@uvt.nl

^dCEB, now Gartner. Email: Luc.Bissonnette@ecn.ulaval.ca

1 Introduction

Population aging has led to a debate on the labor force participation of older workers, increasing the need for understanding retirement decisions and their sensitivity for policy measures. Earlier studies often only focus on the retirement behavior of males, due to the fact that women often left the labor market long before reaching retirement age. Since the labor force participation rate of women has grown substantially over the last decades, understanding their retirement decisions has gained importance. Moreover, there is clear evidence that the retirement decisions within couples are interdependent, with a significant proportion of spouses retiring at approximately the same time, irrespective of their age difference (“joint retirement”). [Hurd \(1989\)](#) already pointed out that between 20 and 30 percent of all couples retire within one year of each other. Since then, a new line of research has emerged aimed at analyzing and understanding the retirement behavior of couples. Empirical evidence of joint retirement decisions has been documented for several cohorts and countries, including [Blau \(1998\)](#) (Retirement History Study), [Gustman and Steinmeier \(2000\)](#) (National Longitudinal Survey of Mature Women), [Gustman and Steinmeier \(2004\)](#) and [Casanova Rivas \(2010\)](#) (Health and Retirement Study), [Banks et al. \(2007\)](#) (English Longitudinal Study of Ageing), and [Hospido and Zamarro \(2014\)](#) (Survey of Health Ageing and Retirement in Europe).

There are several potential links between spouses’ retirement decisions that may explain joint retirement. The first one operates through the household budget constraint (see [Blau \(1998\)](#); [Casanova Rivas \(2010\)](#)). [Casanova Rivas \(2010\)](#) shows that under certain conditions, the fact that resources are shared can increase the probability of joint retirement. In addition, the spouse allowance in the US old age social security benefit system creates financial incentives that explain coordination in retirement dates (see [McCarty \(1990\)](#)). The second channel originates from preferences: It is plausible that spouses enjoy spending leisure time together so that the marginal utility of retirement increases when the partner is also retired

(complementarities in leisure, leading to a causal effect of the retirement of one spouse on retirement of the other spouse; see [Gustman and Steinmeier \(2000, 2004, 2009\)](#), [Stancanelli and van Soest \(2016\)](#) or [Michaud and Vermeulen \(2011\)](#)). The third channel is the correlation in spouses' (unobserved) tastes for leisure (See [Gustman and Steinmeier \(2000, 2004\)](#)) that can be due to assortative matching. The literature has not come to a clear conclusion on the relative importance of these three channels, but does conclude that ignoring joint retirement may severely underestimate the overall impact of reforms in retirement policy. For example, [Coile \(2004\)](#) finds that neglecting the spill-over effect on the behavior of the spouse may underestimate the overall impact of a typical policy by 13% to 20%.

To model joint behavior, the collective model empirically outperforms the neo-classical unitary model assuming that each household behaves as a single decision maker, both for consumption and labor supply behavior ([Fortin and Lacroix \(1997\)](#), [Bourguignon et al. \(1993\)](#), [Cherchye et al. \(2009\)](#)). The collective model ([Chiappori \(1988\)](#)) starts from the basic assumption that in a multi-person household, household members have their own utility functions but cooperate in some bargaining process that results in a Pareto-efficient allocation. The assumption of cooperation seems quite natural, given that spouses interact very often.¹ Several studies found support for the collective model in the sense that its theoretical implications cannot be rejected when tested on multi-person household data ([Chiappori et al. \(2002\)](#), [Cherchye et al. \(2009\)](#)). Yet, some assumptions of that model remain untested. In particular, it is assumed that each spouse knows the preferences of the other spouse and therefore can correctly pick a Pareto optimal outcome. To the best of our knowledge, no study has tested this assumption.

A major challenge of the collective model is the identification of parameters in a realistic context with externalities and public goods. Additional identifying assumptions are usually

¹[Gustman and Steinmeier \(2000\)](#) and several follow-up studies estimate a structural retirement model of joint retirement assuming noncooperative behavior and Nash bargaining.

needed in order to estimate the model and test its validity. [Chiappori \(1988, 1992\)](#) shows that if preferences are egoistic, the usual data on actual household choices are enough to identify the sharing rule up to a constant. The assumption of egoistic preference, however, is rather restrictive in the context of retirement, since it means that the marginal rate of substitution (MRS) between own leisure and consumption remains the same regardless of the spouse's labor supply and therefore does not allow for complementarities (externalities) of leisure activities - one of the possible explanations for joint retirement. Many activities such as taking meals or traveling on holidays will usually be more enjoyable when they can be done together with the spouse. A natural assumption for older couples is therefore that the marginal utility of own leisure depends on leisure of the spouse, that is, preferences are interdependent (See [Gustman and Steinmeier \(2000, 2004, 2009\)](#); [Michaud and Vermeulen \(2011\)](#)). [Michaud and Vermeulen \(2011\)](#) allow for complementarities in leisure, identifying the model by making use of panel data with couples and individuals who became a widow(er) in the observation period, along with the assumption that an individual's preferences can only change after the death of the spouse.

Models of joint retirement behavior also typically assume that expectations regarding the future are rational ([Blau, 1998](#); [Gustman and Steinmeier, 2000](#)). A large literature has studied how expectations regarding mortality, health and other risks deviate from rational expectations in the general population and these deviations are predictive of behavior ([Hurd, 2009](#)). [Manski \(2004\)](#) argues that direct measurement of expectations and their inclusion in behavioral models is desirable in order to make inferences about preferences. Yet, we know of no application of collective models which uses subjective expectations on risks (longevity, disability, etc).

The main contribution of the current paper is to provide a novel approach to obtain identification of a general version of the collective model for labor supply and retirement in a context that allows for subjective (and not necessarily rational) expectations and imperfect knowledge

of preferences of the spouse. Specific to this model is that it avoids imposing restrictions on individual preferences such as egoistic preferences: externalities and public goods are both allowed to enter the individual utility functions. The paper solves the identification problem by using stated preferences (SP) data, aimed at directly identifying the main components of the model. Survey respondents were offered several pairs of simplified retirement trajectories. They were asked to choose between two trajectories three times: first only accounting for their own preferences, then using only their spouse's preferences, and finally they were asked what would be the most likely choice of their household, accounting for both individuals' preferences as well as the decision-making process in their household. As in [Kapteyn and Kooreman \(1992\)](#), the rich SP data directly helps to estimate the individual utility functions and demand for leisure equations that can be used to identify all preference parameters. The answers to the question on the most likely choice in the household then identifies the household decision process and, in particular, the weights of both partner's utility functions for the household choice.

The method of Stated Preferences has been commonly used in marketing and transportation sciences for many years (see, e.g., [Louviere et al. \(2000\)](#)), and is gaining ground in economics since [Barsky et al. \(1997\)](#) and [Revelt and Train \(1998\)](#). [Elsayed et al. \(2018\)](#) and [Van Soest and Vonkova \(2014\)](#) apply stated preferences to labor supply and retirement decisions of individuals. The latter motivate the use of SP data from the fact that they want to analyze preferences for plans that do not yet exist (e.g., retirement after the mandatory retirement age). Moreover, it is often not clear which alternative retirement scenarios workers can choose in real life, when not only the rules of their pension plan but also limited flexibility of their employer may restrict their actual choice set in ways that remain unobserved in the data. Moreover, the actual choice set depends on labor market history and is therefore potentially endogenous to the individual's labor supply preferences. Stated preference data allow for a design where choice opportunities are exactly known, and variation in choices is

substantial and by construction exogenous to preferences. In the current study, an additional strong argument for using SP data is that, as discussed above, with revealed preferences only, not all the parameters of interest can be identified.

The SP data are collected for a subsample of the Health and Retirement Study (HRS). They are used to estimate a stylized structural model of labor supply and retirement of both partners. Our main findings are the following. We find that males have a biased view of their wives' labour supply preferences, overestimating their disutility of work. We correct for this bias in the model. We find significant complementarities in leisure and unobserved heterogeneity in determining joint retirement. The difference between the male and female's wage rate significantly affects the bargaining weights, confirming that the unitary model is rejected against the collective model. Simulations based upon the estimates (that corrects for the biased perception of partner preferences) suggest that correlation in unobserved heterogeneity components of the two partner's marginal utility of leisure can explain a large share (almost half) of joint retirement decisions. We also find significant positive complementarities in leisure, but this explains a much smaller part of joint retirement.

The remainder of this paper is organized as follows. Section 2 uses a simple example to illustrate the idea of identification in a stylized collective model. Section 3 describes the sample and our SP data. The empirical specification of the structural collective model is discussed in Section 4. The estimation results are discussed in Section 5. Section 6 presents the results of some simulation exercises, validating the SP data by comparing simulated retirement ages with data on expected retirement ages of the same individuals, and then using counterfactual simulations to indicate how complementarities in leisure and correlation between preferences of the two partners contribute to joint retirement outcomes. Section 7 concludes.

2 Identification: A Simple Illustration

We consider a simple example of a static (one time period) collective labor supply model for couples consisting of a man (m) and a woman (f). For $j = m, f$, l_j denotes individual j 's leisure. We work with one aggregate consumption good with price normalized to 1; its amount is denoted by c . Consumption is assumed to be public, in the sense that it gives utility to both spouses. The time endowment for each individual is denoted by T . The wage rates are w_m and w_f and y is the household's non-labor income. Y represents full household income given by $Y = (w_m + w_f)T + y$. In the general framework, individual j 's preferences are represented by a utility function $U_j(l_m, l_f, c; z)$ where z is a vector of taste shifters.

Under the assumption of the collective model that the couple reaches a Pareto-efficient equilibrium ([Chiappori \(1988\)](#)), there exists a weighting factor λ such that (l_m, l_f, c) solves the following maximization problem:

$$\max_{(l_m, l_f, c)} \lambda U_m + (1 - \lambda) U_f \tag{1}$$

subject to the budget constraint

$$c + w_m l_m + w_f l_f \leq Y \tag{2}$$

The weight λ can be a function of w_m, w_f, y and other variables d (called the “distribution factors”).

This general collective model cannot be identified from the usual data on labor supply of both partners and consumption. Identification relies on additional assumptions such as some type of separability restrictions (see [Chiappori \(1988, 1992\)](#)). For example, identification can be achieved if preferences are assumed to depend on consumption and own leisure only, and not on leisure of the partner. As we discussed before, such egoistic preferences are restrictive,

especially for older couples. In this paper, we will therefore not impose such an assumption and allow for interdependent preferences instead of imposing egoistic preferences.

More generally, we avoid any additional assumptions but instead collect additional data to identify the model. To illustrate this idea, we will use a simple example. The arguments of each utility function are $q = (l_m, l_f, c)$ and we normalize T to 1. We assume that preferences can be represented by a Stone-Geary type direct utility function, with constant returns to scale and zero subsistence levels:

$$u_m(q) = \alpha_m \ln l_m + \alpha_f \ln l_f + (1 - \alpha_m - \alpha_f) \ln c \quad (3)$$

$$u_f(q) = \beta_m \ln l_m + \beta_f \ln l_f + (1 - \beta_m - \beta_f) \ln c \quad (4)$$

The collective model implies that the household bargaining process results in a Pareto efficient allocation, solving

$$\max_q u_j(q) = \lambda u_m(q) + (1 - \lambda) u_f(q) \quad (5)$$

subject to the budget constraint

$$c + w_m l_m + w_f l_f \leq Y \quad (6)$$

It is straightforward to show that this gives the following demands for leisure:

$$w_m l_m = (\lambda \alpha_m + (1 - \lambda) \beta_m) Y \quad (7)$$

$$w_f l_f = (\lambda \alpha_f + (1 - \lambda) \beta_f) Y \quad (8)$$

Let us consider the case where the bargaining weights do not depend on wage rates or

non-labor income y , and the weight of the husband is given by

$$\lambda = \gamma_0 + d'\gamma \quad (9)$$

The equations for male and female leisure can then be written as:

$$w_m l_m = (d'\gamma(\alpha_m - \beta_m) + \gamma_0 + \beta_m)Y \quad (10)$$

$$w_f l_f = (d'\gamma(\alpha_f - \beta_f) + \gamma_0 + \beta_f)Y \quad (11)$$

It is easy to see that the parameters in these equations are not identified if only actual behavior $(c, l_m, l_f, w_m, w_f, z)$ is observed – It is necessary to fix (for example) γ_0 and the scale of the vector γ . [Kapteyn and Kooreman \(1992\)](#) emphasize the same problem in a static labor supply model. They propose to use stated preferences on desired working hours by each individual to identify all the parameters. Our approach will be similar in spirit – we collect and use stated preference data that help identification. In our example, suppose respondents report how many hours respondents would like to work themselves and how many hours they would like their partner to work, accounting for their own preferences only and not for those of their spouse. For males, this would be the outcome of the collective model in the case of male dictatorship ($\lambda = 1$); for females, it would correspond to female dictatorship ($\lambda = 0$). The information provided by male respondents can be used to estimate the parameters α_m and α_f of the male utility function, using the two leisure demand equations

$$w_m l_m = \alpha_m Y \quad (12)$$

$$w_f l_f = \alpha_f Y \quad (13)$$

Similarly, the information for female respondents can be used to estimate the parameters

β_m and β_f of the female utility function, using the equations

$$w_m l_m = \beta_m Y \tag{14}$$

$$w_f l_f = \beta_f Y \tag{15}$$

Finally, if we have data on actual labor supply (as in [Kooreman and Kapteyn \(1990\)](#)) or ask the respondents to also report the most likely actual outcome for the household, it is easy to see from (10) and (11) that the bargaining parameters (γ_0, γ) are also identified. In particular, this also identifies the intercept γ_0 which is not identified in the collective models of [Chiappori \(1988, 1992\)](#); [Chiappori et al. \(2002\)](#). This intercept is needed to interpret the estimation results in terms of the bargaining power of each partner.

The example shows that stated preferences on what individuals in couples would do if they were the only decision makers, that is, in case of male or female dictatorship, lead to information to identify parameters in the collective model that cannot be identified by revealed preferences alone. In this paper, we exploit this idea in the context of an inter-temporal (multiple time periods) model explaining retirement decisions. We will use information on survey respondents' stated choices among different hypothetical joint retirement trajectories (retirement ages and corresponding earnings and pension income for both partners) based on their own preferences and based upon their spouses' preferences. To identify the bargaining weight, we will then also use the answers to the question on the most likely outcome for the couple.

3 Data

Our module of SP questions was included in the 2011 Internet Survey (2011-IS)² of the Health and Retirement Study (HRS). The HRS is an ongoing longitudinal biennial socio-economic survey of the U.S. population aged 50 years and older and their spouses, conducted by the Institute for Survey Research at the University of Michigan. The 2011-IS is developed jointly by the HRS, the Survey Research Center (SRC) and the Institute for Social Research (ISR) at the University of Michigan, and the RAND Corporation. It is the fifth in a series of surveys administered over the Internet to a sub-sample of HRS respondents with Internet access in years where the HRS respondents do not get the regular HRS interviews. Over the years, the HRS Internet has had many experimental question modules covering a large variety of topics including, for example, subjective probabilities (Delavande and Rohwedder, 2008), early childhood health histories (Smith, 2011), health literacy (Levy et al., 2015), and household financial assets (Couper et al., 2013).

In the subsample selected for the questions on joint retirement, we only selected respondents in couples who were doing paid work themselves or whose partner did paid work. The number of respondents participating was 779, but we only use the observations where both respondent and spouse were younger than 62. This is because scenarios were phrased differently in other cases and because we want to use expected work limiting health problems as a covariate. Moreover, we removed some observations with missing information on important variables (subjective survival probabilities, whether the respondent or their partner has a work limiting health problem, the probability that they will have such a problem at age 62, job satisfaction; missing wage rates were imputed). This leaves a final sample of 604 couples, with 280 male and 324 female respondents.

Table 1

²Link: <http://hrsonline.isr.umich.edu/index.php?p=shownews3x1&hfyle=news319>

Summary statistics of the explanatory variables for the male and female partners in the households of the 604 survey respondents in our final sample are given in Table 1. The average ages of males and females are 56.4 and 54.5 years, respectively. The age difference between spouses has a standard deviation of 4.72 and an interquartile range of 4 years. Almost the same fraction of males and females have a college degree (44% of men and 42% of women). Satisfaction with the current job is included in the model since it may affect the marginal utility of leisure (as opposed to working). The fraction of men and women who are satisfied (or very satisfied) with their job is almost the same (75%). Males somewhat less often report to have a work limiting health problem (7% versus 9%). On the other hand, the average subjective probability of a work limitation at age 62 is very similar for both genders (40% and 39%). Since subjective survival probabilities until age 75 depend on current age, Table 1 does not present the reported probabilities themselves, but their ratio to the (gender specific) objective survival probabilities according to Human Mortality Database (www.mortality.org).

Males earn higher hourly wages than females. Accordingly, the log of the wage ratio (male's wage rate divided by female's wage rate), which is used in the model as a determinant of within household bargaining power, has a positive mean of 0.242 and substantial variation around this mean, with a standard deviation 0.753 and an interquartile range of 0.908.

We designed six hypothetical and stylized retirement trajectories. All the scenarios are in terms of income-hours combinations for both partners over the remaining part of the life cycle after the respondent has reached a given age. All scenarios are designed to be realistic, in the sense that they are broadly similar to the choice opportunities individuals and couples can encounter in real life. Respondents were explicitly asked to assume that their employers will cooperate with each scenario.³

Figure 1

³For the complete questionnaire, see pp. 391–417 of <http://hrsonline.isr.umich.edu/modules/meta/2011/internet/qnaire/online/net2011qnaire.pdf>

Figure 1 gives a screen shot of an example of a retirement trajectory of the couple, i.e., both for the respondent (top panel) and the respondent’s partner (bottom panel). The respondent’s remaining life time from age 62 was divided into three periods: 62-65, 65-68, and beyond 68. This remains the same for all scenarios; these ages were chosen since transition rates into retirement tend to peak at these ages. For the partner, we use the same division in calendar time. In Figure 1, the partner is two years younger than the respondent, so if the respondent is, for example, in the age bracket 65–68, then the partner is (roughly) in the age bracket 63–66.

Initial hours of work were always set to the respondent’s and partner’s actual working hours before (partial) retirement (40 hours per week for both respondent and partner in Figure 1). Once they are fully retired, individuals get a pension corresponding to a proportion of their last earnings (Replacement Rate (RR)). During partial retirement, their disposable income is a combination of earnings and pension. On the screen, we present the corresponding total replacement rate (disposable income during partial or full retirement as a fraction of pre-retirement disposable income). In the example of Figure 1, the respondent retires partially at age 65, reducing working hours from 40 to 24 per week, with a total replacement rate of 100% (earnings plus partial pension). The respondent fully retires at age 68, after which the replacement rate is 75%. The respondent’s partner retires at age 63 (when the respondent turns 65), with a replacement rate of 75%. There is no partial retirement for the partner.

Variation in replacement rates was created by randomly varying them across the sample. More precisely, respondents are divided into five groups indicated by G_1, G_2, G_3, G_4 and G_5 . Group G_5 got the highest RRs in all scenarios, group G_4 got the second highest RRs, etc. Since the design of the scenarios takes account of the respondent’s actual age as well as the age difference, couples with different actual ages and a different age difference will receive different retirement with adjusted replacement rates. In Table 2, we illustrate the different scenarios, using the example of a male respondent and his spouse who are both 55 years old

and who both work 40 hours per week. The table shows the retirement plans presented to the respondent in the survey, with its (partial and full) retirement ages and the replacement rates for the five different groups. Scenarios 1, 2, 3 and 6 have abrupt retirement of both partners. In Scenario 1, the male respondent retires earlier (age 65) than the wife (age 68), in scenario 2 it is the other way around (ages 62 and 68). Scenario 4 has partial retirement of the respondent, who works part-time from age 62 until 65. Working part-time always means working 60% of preretirement hours. In scenario 5, the partner works part-time from age 65 until age 68. During partial retirement, the replacement rate is always 100%; after full retirement, replacement rates are randomly assigned through the group assignment G1–G5, as shown in the table. If, unlike in the example of Table 2, the age difference is not zero, the (partial) retirement ages and RRs for the partner change accordingly.

Table 2

In order to stimulate respondents to think of the pros and cons of each scenario before asking them to make choices, respondents were first asked to rate every scenario separately on an ordinal scale. These ratings will not be used in the empirical analysis; they merely serve to familiarize the respondents with all the scenarios. The reason is that in a similar SP experiment for individual retirement trajectories, [Van Soest and Vonkova \(2014\)](#) used both rating and choice data but found that stated ratings were very noisy, giving much less information than the stated choices. Choice questions come after these ratings. Respondents were asked to indicate a preference between the two scenarios in each of three pairs: scenario 2 versus scenario 4, scenario 5 versus scenario 2, and scenario 3 versus scenario 6. For each of these three choice sets comparing two scenarios, three questions were asked:

I: Considering your own preferences and not those of your partner, which would you prefer?

II: Considering your partner's preferences and not those of yourself, what do you think your partner would prefer?

III: If, as a couple, you would have to decide jointly on choosing between these scenarios, what would be the most likely outcome?

In total, we have three pairs with three choices each. As we previously mentioned, the three different types of questions are important for identification of the collective model. Questions *I* and *II* allow us to elicit the preference parameters of males' and females' individual utility functions, while question *III* is informative on the parameters driving the bargaining weights of husband and wife.

Table 3

Table 3 presents descriptive statistics for the nine stated choice questions. Each row represents one question; each of the three blocks has the three questions for a given choice set. In the first block where the choice is between scenarios S2 and S4, both husbands and wives tend to prefer S4 over S2 from all three perspectives. S2 was chosen more often by female than by male respondents. From Table 2 we know that S4 focuses on gradual retirement of the respondent while S2 has early retirement, while there is no difference for the partner. Female respondents more often than male respondents to stop working completely at age 62 than to continue working part-time, even if not working at all leads to less income. This is in line with the notion that women have a lower preference to do paid work. The same explanation can be given for the difference between rows 1 and 2: A larger preference is given to early retirement when respondents answer the question using the wife's preferences than when using the husband's preferences. As expected, the choice fractions from the perspective of the household as a whole are generally in between the fractions from the male's and female's perspective.

For the second choice situation, both male and female respondents tend to choose S5 over S2. Again, females tend to prefer S2 more often than males, in line with the notion that females tend to have lower labour supply than men. Note that both S4 and S5 have partial retirement for either the respondent (S4) or the spouse (S5). A preference for S4 or S5 therefore also implies a preference for partial retirement, in line with what [Van Soest and Vonkova \(2014\)](#) and [Elsayed et al. \(2018\)](#) found using Dutch SP data and analyzing individual preferences. For example, [Van Soest and Vonkova \(2014\)](#) analyzed four types of hypothetical retirement scenarios - standard retirement, late retirement, early retirement and partial retirement and found that if more flexibility is added in the form of gradual retirement with an actuarially fair partial pension, almost two thirds would opt for gradual retirement.

For the last SP question, females and particularly males tend to choose S6 (standard retirement for both) over S3 (late retirement for both) from all three perspectives. As expected given the previous results, late retirement is particularly unpopular among female respondents. Looking at the table as a whole, we can conclude that both male and female respondents tend to make similar choices regardless of the perspectives they are supposed to choose from – the differences across choice situations are much larger than the differences across perspectives for a given choice situation.

4 Empirical specification

As described in Section 3, the respondents need to choose between different retirement trajectories for the couple. We assume that for a respondent $j = m, f$ in couple i , total lifetime utility U_{iq}^j of the joint retirement trajectory q has the following additively separable form:

$$U_{iq}^j = \sum_{d=0}^1 p_d \sum_{t=62}^{100} \rho^{t-62} \Pr(t|62) U_{iqt,d}^j \quad (16)$$

Here $U_{igt,d}^j$ is the utility at age t with work limitation status $d \in \{0,1\}$, $\Pr(t|62)$ is the probability of survival till age t given survival until age 62, p_d is the probability of having work limitations after age 62,⁴ and ρ is the discount factor. Both health and survival risk are taken from respondents' responses. The survival probabilities are constructed adjusting general life tables for men and women upward or downward, using subjective probabilities of living until age 75 reported in the HRS for the respondent and the spouse.⁵ The summation is from age 62 for the respondent since all of our SP scenarios start from respondent age 62. For the spouse, we use the same intertemporally additive specification, but the summation is adjusted for the age difference between respondent and spouse, in line with how we designed the scenarios. For example, if the partner is two years younger than the respondent, the summation for the partner is from partner's age 60.

Since scenarios are presented in terms of income rather than expenditures, we also work within period utility functions in terms of income rather than expenditure. In other words, we abstract from saving and consumption smoothing, as in [Van Soest and Vonkova \(2014\)](#). The within period utility functions U_{igt}^j are assumed to have the following modified Cobb-Douglas specifications:

$$U_{igt}^m = \alpha_i^m \ln l_{igt}^m + \alpha^f \ln l_{igt}^f + \alpha^c \ln c_{igt} + \alpha^{mf} \ln l_{igt}^m \ln l_{igt}^f \quad (17)$$

$$U_{igt}^f = \beta^m \ln l_{igt}^m + \beta_i^f \ln l_{igt}^f + \beta^c \ln c_{igt} + \beta^{mf} \ln l_{igt}^m \ln l_{igt}^f \quad (18)$$

$$\alpha_i^m = x_i^m \psi^\alpha + \eta_i^m \quad (19)$$

$$\beta_i^f = x_i^f \psi^\beta + \nu_i^f \quad (20)$$

⁴Ideally, this probability should depend on t but the data only contain the probability at age 62.

⁵Let the objective life table survival rate to age t be $s_o(t) = \exp(-\Lambda_o(t))$ where $\Lambda_o(t)$ is the integrated hazard and we assume the subjective survival rate to age t , $s_s(t) = \exp(-\psi\Lambda_o(t))$. We use both subjective and objective point estimates to find ψ for each respondent. We then reconstruct the subjective survival curve using the integrated hazard of the life-table survival curve.

Here l_{iqt}^j and c_{iqt} denote partner j 's leisure and household (public) consumption, respectively. Utility depends on own leisure, the spouse's leisure, consumption and the interaction of the two leisure terms. x_i^j are vectors of taste shifters, capturing observed heterogeneity (age, health indicators, education, etc). α_i^m and β_i^f are linear combinations of observed and unobserved heterogeneity terms; η_i^m and ν_i^f are the unobserved heterogeneity terms, which vary across households but are fixed over time. We use a bivariate normal distribution of (η_i^m, ν_i^f) with parameters to be estimated, allowing the two random coefficients to be correlated. A positive correlation would capture the fact that men and women with similar preferences for leisure are more likely to become partners ("assortative matching" on the marriage market). All other preference parameters $(\alpha_c, \beta_c, \alpha_{mf}, \beta_{mf}, \rho)$ are assumed to be the same for all couples in the sample.

It is clear that this specification allows for externalities in leisure: utility depends upon leisure of the partner. In particular, the interaction terms imply that marginal utility of own leisure varies with leisure of the partner. If there are complementarities in leisure, α^{mf} and β^{mf} will be positive – implying that the marginal utility of leisure increases with leisure of spouse.

The bargaining weight of males is given by $\lambda(p; \delta)$. We use the following logistic transformation to guarantee that the weight is bounded between 0 and 1:

$$\lambda(p; \delta) = \frac{\exp(p'\delta)}{1 + \exp(p'\delta)} \quad (21)$$

For simplicity, we assume that p only includes the log wage ratio $(\ln(w_m/w_f))$ and an intercept. We expect a positive effect of $\ln(w_m/w_f)$ on λ since this implies that a higher relative wage of the husband compared to the wife gives the husband a better bargaining position. The intercept is also of interest, since it reflects the bargaining position if both partners have the same wage rate.

To account for optimization errors, general extreme value Type 1 (GEV(1)) iid error terms ϵ_{iq}^m , ϵ_{iq}^f and ϵ_{iq}^j are added. For questions of type *I* for males and type *II* for females in every choice question, the scenario q is chosen that maximizes:

$$V_{iq}^m = U_{iq}^m + \epsilon_{iq}^m \quad (22)$$

Similarly, for questions of type *I* for females and of type *II* for males, the scenario q is chosen that maximizes:

$$V_{iq}^f = U_{iq}^f + \epsilon_{iq}^f \quad (23)$$

For questions of type *III*, respondents choose the retirement scenario q that maximizes:

$$V_{iq}^j = \lambda U_{iq}^m + (1 - \lambda)U_{iq}^f + \epsilon_{iq}^j \quad (24)$$

Given the specification of α_i^m and β_i^f , our model can be estimated as a random coefficient model. Because optimization errors are independent of each other, the conditional likelihood contribution of respondent i given unobserved (and observed) heterogeneity terms can be written as a product of the probabilities P_{is} of the observed choices in questions $s = 1, \dots, S$. Each of these probabilities is a simple binary logit probability.

The parameters determining the distribution of the unobserved heterogeneity terms (the K mass points and their probabilities p_k) are estimated jointly with the other parameters using maximum simulated likelihood.⁶ The exact (unconditional) likelihood to be maximized has the form:

⁶Parameters that are transformations of the probabilities are estimated to guarantee that probabilities are positive and add up to unity.

$$L = \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{s=1}^S P_{is}(\eta_i^m, \nu_i^f) f_{\Sigma}(\eta_i^m, \nu_i^f) d(\eta_i^m, \nu_i^f) \quad (25)$$

Here $f_{\Sigma}(\eta_i^m, \nu_i^f)$ denotes the bivariate normal density of the unobserved heterogeneity terms (η_i^m, ν_i^f) , characterized by a covariance matrix Σ (and mean $(0,0)$). This likelihood is approximated, replacing, for each couple i , the integral, which is an expectation over (η_i^m, ν_i^f) , by a simulated mean of R draws.⁷

5 Estimation Results

Table 4 gives the estimates of our benchmark collective retirement model. The first part presents the influence of taste shifters on the marginal utility of an individual's leisure. The marginal utility of leisure increases with age, in line with the idea that keeping taste shifters and financial incentives constant, individuals prefer to reduce work effort when they get older. This is consistent with existing findings, such as [Michaud and Vermeulen \(2011\)](#).

Table 4

We tried several health indicators (cf. Table 1), and the best fit was obtained with the dummy for health limitations. The significantly positive estimates indicate that individuals with health problems have a higher disutility of paid work than those without health problems, as we would expect. We do not find significant differences in the marginal disutility of work with education or satisfaction with the current job. The positive (but insignificant) estimates for men are somewhat surprising, since they would imply a higher disutility of work for men who are satisfied with their job.

⁷following standard practice, these draws are transformed standard normal draws, using the Cholesky decomposition of Σ . We use Halton draws to reduce simulation variance and use $R = 50$; see, e.g., [Train \(2009\)](#).

The final “taste shifter” we included is denoted as “spouse responded”. For men, it is equal to 1 if the respondent is the wife; for women it is equal to 1 if the husband responded to the survey. If the respondent was perfectly aware of the partner’s preferences, the parameter on this covariate would be zero. Point estimates are positive, insignificant for men, but significant for women. This provides evidence that on average, men have a biased view of the preferences of their wives. They overestimate their wives’ marginal utility of leisure, and will therefore more often choose scenarios with earlier retirement of the wife than their wives themselves choose. In the simulations we perform in the next sections, we will take out this “information bias”. This is all based on the assumption that individuals know their own preferences perfectly well, so that inconsistencies between their own and their partner’s reported preferences must be due to a biased perception of partner preferences.

One of the main features of this model is that preferences are not egoistic and we allow externalities in leisure, reflected in the parameters α^{mf} and β^{mf} in equations 17 and 18, the parameters on “leisure spouse” in Table 4. Both α^{mf} and β^{mf} are positive and significant, confirming the notion that joint leisure activities imply that the marginal utility of own leisure increases if the spouse also has more leisure. Similar evidence can be found in, for example, [Gustman and Steinmeier \(2000\)](#) and [Michaud and Vermeulen \(2011\)](#). This result also suggests that the assumption of egoistic preferences, ignoring possible externalities in leisure, may lead to incorrect results. This problem may of course be specific to the context of retirement and older couples, it is not necessarily an issue for labour supply of prime age men and women. In the next section, we will use simulations based upon our estimates to illustrate how much of joint retirement in the data is explained by the estimated importance of complementarity of leisure.

The marginal utility of consumption is positive and significant for men as well as women (parameters α^c and β^c). The marginal utility of leisure (also accounting for the unobserved heterogeneity term) is positive in the majority of cases for both males and females, but not

always. Particularly at younger ages, a substantial fraction of men and women have negative marginal utility of leisure, reflecting the fact that there is a group of respondents who never choose early retirement.

The parameter δ_2 for the log wage ratio in the bargaining weight λ (equation 21 in Section 4) is marginally significant (t-value 1.67), implying that a higher wage rate relative to the wage rate of the partner increases the individual’s Pareto weight and bargaining power. This also implies that the standard unitary model (implying that the household utility function in equation 24 should be independent of the wage rates) is (marginally) rejected against the more general collective alternative. This is in line with most findings in the collective model literature. In our setting, we can also identify the constant δ_1 . The estimate is positive but very small and not significant. This means that if both spouses earn equal wages, they tend to have almost the same bargaining power (0.503 for the husband, 0.497 for the wife). If the wage rate of the husband is twice that of the wife, the estimates of δ_1 and δ_2 imply that the husband’s Pareto bargaining weight is approximately 0.592.

Both discount factors are estimated to be larger than 1. This is much larger than what is found in studies that do not explicitly incorporate survival probabilities, but similar to the discount rates found by French (2005). Like we do, he separately incorporates mortality risk and finds discount rate estimates varying from 0.981 to 1.04, suggesting much more patience than typically assumed or found when ignoring mortality.

The model has two unobserved heterogeneity terms, the random components of the coefficients α^m and β^f . Both are significant, and they are almost perfectly correlated, with an estimated correlation coefficient of 0.995, not significantly different from 1.⁸ This implies a strong positive correlation between husband’s and wife’s preferences for leisure. One plausible

⁸The covariance matrix is estimated as a function of three auxiliary parameters using a Cholesky decomposition, modelling ν_i^f in 20 as a linear combination of η_i^m in 19 and a component independent of η_i^m . The parameter $\sigma(\eta_i^m)$ and the parameter determining how much ν_i^f is related to η_i^m are both strongly significant, but the parameter driving the independent component of ν_i^f is small and insignificant (t-value 0.5).

explanation is assortative matching on the marriage market (men and women who get married tend to have similar preferences), another possibility is that the preferences of husbands and wives in these older couples (one of the two is at least age 50) have converged over time. In the simulation exercises of the next section, we analyze to which extent this correlation can explain joint retirement.

We estimated several alternative specifications that are less general than the baseline model presented in Table 4. Table 5 shows how these alternative specifications compare to the baseline in terms of log likelihood. Since each of the alternative models is nested in the baseline model, they can also be compared to the baseline model using a formal likelihood ratio test on one or two parameter restrictions. The LR test statistic is presented in the second column of the table. This shows that all the alternatives are rejected against the more general baseline. Even the unitary model is clearly rejected against the baseline model, with a p-value of 0.018; for the other models, the p-values are much smaller still.

Table 5

Detailed results for the other models can be found in the appendix. In general we find that, even if the restrictions imposed in the alternative models are clearly rejected by the tests, imposing these restrictions does not have large effects on the estimates of the other parameters. For example, the correlation between unobserved heterogeneity terms is always close to one (except when restricted to zero) and the discount factors are always above 1 (except when set to 0.95). The coefficient on the log wage rate ratio varies somewhat more, from 0.391 to 0.533, with t-value between 1.33 and 1.84.

6 Simulations: Validation, explaining joint retirement, and financial incentives

The parameter estimates provide strong evidence of both complementarity of leisure and correlation between labour supply preferences of husband and wife. The main purpose of this section is to perform some simulations to illustrate how much complementarity of leisure and correlated preferences contribute to explaining joint retirement. The simulation setup is similar to the policy simulations in [Van Soest and Vonkova \(2014\)](#), but for couples instead of individuals, and focusing on explaining joint retirement rather than gradual retirement or measuring the effects of financial incentives. The simulations let households choose among stylized retirement scenarios similar to the ones presented in our stated preference questions. This implies that we do not consider a third possible explanation of joint retirement: joint features in the system of taxes and benefits. In our simulation set up, household income in each time period is simply the sum of the husband’s and the wife’s earnings and old age pensions, without any joint feature like the spouse allowance or health insurance through the partner’s employment.

In the simulations, we do not consider gradual or partial retirement and thus allow for abrupt retirement only. Moreover, retirement is treated as an absorbing state, there is no reverse retirement. We assume males and females can both retire at any age from 60 to 70. This gives 121 potential outcomes for the household. In the baseline simulation, the replacement rate for each individual is set to 60% if the individual retires at age 65. It is adjusted upward with 8% per year if the individual retires later (a “delayed retirement credit” of 8% per year) and reduced by 6% per year if the individual retires earlier (an “actuarial reduction factor” of 0.94).⁹

⁹These adjustments make the reward for later retirement and the penalty for earlier retirement roughly actuarially fair.

Our model implies that the probability of choosing a given cell depends on observed taste shifters, wage rates, estimated parameters, and the unobserved heterogeneity terms. Taste shifters and wage rates are taken from the data and parameter estimates are taken from Table 4. Each couples unobserved heterogeneity components η_i^m and ν_i^f are estimated as the posterior means, updating the estimated bivariate prior distribution with the observed choices in the data. For each couple, we then use these posterior means and all the other information to compute expected retirement ages of both partners.

Figure 2

A contour plot of the resulting distribution of retirement ages is sketched in Figure 2. The modal cell is approximately age 62 for both, but the distribution is spread out and strongly right skewed, particularly for males (whose marginal utility of leisure is, according to the estimates, characterized by substantially more unobserved heterogeneity than the marginal utility of leisure for women). The contour plot also shows a positive correlation between the retirement ages of both spouses.

External validation

According to the baseline simulation with the given, stylized, choice opportunities, the average retirement age is 65.73 for men and 63.83 for women. It is hard to compare these numbers with average retirement ages in actual data, since real life choice opportunities will almost certainly deviate from the 121 points stylized choice sets used in the simulations. To validate our stated preference data, we can, however, compare our expected retirement ages for each couple to the expected retirement ages reported by the same individuals in the HRS (wave 2010; the answer to the question “when do you think you will stop working?”).

Figure 3 presents the results We find a clear positive and significant relation between the simulated retirement ages according to our model and the expected retirement ages in the

HRS data. This is what we expect if our SP data are valid: individuals with a preference for early retirement retire earlier according to our SP data and the model based upon these data, and also indicate a lower expected retirement age. The R-squared of the linear regression is 0.21 for men and 0.25 for women. This indicates that the correlation is far from perfect, due to, among other things, the fact that the choice sets respondents use may very well differ from the stylized choice sets in the simulations, and noise in both the simulated and the reported expected retirement ages. Still, the fact that our SP data help to predict the expected retirement ages reported in the HRS shows the usefulness of the SP data in addition to revealed preferences, particularly in the context of serious identification issues that we have in the collective model (cf. Section 2).

Explaining joint retirement

Figure 2 does not say much about joint retirement yet, since the two spouses may have different ages – Joint retirement means retirement at (approximately) the same calendar time, not at the same age. Accounting for the age difference between partners, however, can easily compute the difference in calendar time between the expected time of retirement of husband and wife. We refer to this as the retirement *distance*. We will consider the distribution of *distance*, but if we want a percentage of couple who retire jointly, we have to be more specific. We will define joint retirement as the event that *distance* is at most one in absolute value. This is a rather broad definition, allowing one year of difference between the two retirement ages.

The top row of Table 6 shows that with the given choice sets, the model predicts that 33.6% of couples retire jointly according to this definition.¹⁰ Figure 4 illustrates the joint retirement decisions further, presenting a contour plot in terms of the simulated distance to retirement for husbands and wives at the time of the survey. The shaded green area reflects

¹⁰Gustman and Steinmeier (2000), who assumed couples attain a noncooperative Nash equilibrium and predicted that over 11% of couples retire at the same time. Our number cannot be compared to theirs, since they use a completely different framework and definition

joint retirement outcomes, and the figure clearly shows that this area covers a range where the density of simulated retirement ages is relatively high (the distribution is unimodal with the largest density at approximately (10,6); lighter contour lines indicate higher density).

The remaining rows of Table 6 show how the percentage of joint retirement changes in a counterfactual simulation where one of the potential explanations for joint retirement is taken out. The row “Shuffling wages” shows what happens if randomly reassign the males’ and females’ wage rates in the sample to the couples in the sample. This takes out one source of correlation due to observed heterogeneity: males’ and females’ wage rates are now independent of each other (but also independent of all the taste shifters). As expected, average simulated retirement ages of men and women hardly change compared to the baseline simulation. The simulated probability of joint retirement is, somewhat surprisingly, even larger than according to the baseline simulation. This suggests that the positive correlation in husbands’ and wives’ wage rates (0.056 in the data) does not explain joint retirement. This is also illustrated in Figure 5. The blue line in this figure gives the baseline simulation of the distribution of *distance*. The orange line is the distribution according to the counterfactual simulation with wages reshuffled, giving somewhat more joint retirement $-1 \leq \textit{distance} \leq 1$) than the baseline.

The third row in Table 6 eliminates the correlation between the two unobserved heterogeneity terms. Here we use unobserved heterogeneity terms with the same variances as in the baseline simulation, but with zero correlation coefficient. This makes quite a difference, since the estimated correlation used in the baseline simulation is 0.995. Joint retirement drops dramatically, from 33.6% to 17.7% according to our broad definition. In other words, correlated labour supply preferences, due to assortative matching or convergence in preferences over time, can explain almost half of all joint retirement decisions. This is much higher than what is found in [Gustman and Steinmeier \(2000, 2004\)](#).

The final counter-factual simulation in Table 6 focuses on the role of complementarities

in leisure. In this simulation, we fixed the spouse’s leisure in the individual utility function to the average level in the sample, so that the individual’s marginal utility of leisure no longer depends on the partner’s leisure. This simulation gives a somewhat smaller probability of joint retirement: 30.4% instead of 33.6% in Table 7 (broad definition). It implies that complementarity in leisure can explain almost 10% $((33.6-30.2)/33.6)100\%$ of all joint retirement decisions. See also the red curve in Figure 5, which has less density around 0 than the blue baseline curve. This result is similar to that of [Casanova Rivas \(2010\)](#), who, using a stochastic dynamic optimization model, found that leisure complementarities account for 13% of observed joint retirement. Our simulations results therefore lead to the conclusion that unobserved heterogeneity plays an much more important role than complementarities in leisure. [Stancanelli and van Soest \(2016\)](#) use a double regression discontinuity approach to investigate joint leisure before and after retirement. Their results also suggest that leisure complementarities are not the main driver of joint retirement.

Joint retirement and financial incentives

Table 7 illustrates how sensitive individual and joint retirement decisions are for the financial incentives, even if these incentives do not directly create joint features in the budget restrictions. The top row reproduces the baseline simulation already discussed above. The second row shows what happens if there is no actuarial penalty on retiring earlier and no reward for retiring later. This reduces incentives to work and therefore lead to earlier retirement for both partners. Retirement ages will be much more concentrated in the range 60-65, implying that more couples retire in the same year or adjacent years.¹¹ Accordingly, joint retirement (according to our broad definition) increases from 33.6% in the baseline simulation to 43.5% in the “no penalty” simulation. The “high penalty” simulation is essentially the opposite of

¹¹Not everyone retires at the earliest possible age, since wages are higher than pensions and because, according to our estimates, a substantial group of individuals has negative marginal utility of leisure at early retirement ages (partly due to the large unobserved heterogeneity component).

the “no penalty” case. Rewards for later retirement are higher than in the baseline situation (see notes to the table for details), and individuals are stimulated to retire later. The average retirement ages of men and women increases, but the distribution of retirement ages is also more spread out, so that fewer couples retire at approximately the same time. Accordingly, the percentage retiring at most one year from each other drops from 33.6% to 27.6%.

The final two rows concern what is sometimes called the income or pension wealth effect on retirement (cf. [Van Soest and Vonkova \(2014\)](#)). Without changing rewards for earlier or later retirement compared to the baseline, pensions are made less (“low generosity”) or more (“high generosity”), irrespective of the chosen retirement rate, so that total pension wealth and life-cycle income are lower or higher than in the baseline simulation. If leisure is a normal good, lowering life-cycle income (“low generosity”) leads to more labour supply, i.e., later retirement. This is indeed what the table shows for both spouses, although the effects are not very large, considering the substantial reduction in replacement rates (from 60% to 40% when retiring at age 65. Accordingly, like in the “high penalty” simulation, retirement ages are more spread out and joint retirement becomes less likely. The final row (“high generosity”) gives essentially the opposite effect: higher pensions lead to earlier retirement, more concentrated around the early ages. Joint retirement increases from 33.6% to 36.2%.

All in all, the bottom panel of the table shows that apart from complementarities in leisure and correlation in preferences, the characteristics of the choice set (the budget opportunities) are an important determinant of the fraction of couples retiring jointly, even if pension income is not explicitly linked to the retirement age of the partner.

7 Conclusion

The collective model is valuable tool to describe the behaviour of multi-person households, outperforming the unitary model which assumes that the household acts as a single utility-

maximizing agent. On the other hand, estimating a collective model requires a lot from the data, and revealed preference data alone are typically insufficient without additional assumptions. Furthermore, it typically involves making strong assumptions regarding expectations and knowledge of preferences within the couple. In this paper, we consider a collective model of labour supply and retirement of couples that allows for subjective expectations and misperceptions of spouse preferences. We show that additional assumptions can be avoided when stated preference data are used instead of revealed preference data. Asking survey questions on choices between different hypothetical retirement scenarios made from three different perspectives, own preferences, partner preferences, and most likely outcome for the couple, leads to a natural way of identifying preference parameters as well as parameters driving the household decision.

Stated preferences have been successfully applied to study individual retirement decisions. We acknowledge that the stated preference questions for couples are more challenging than those for individuals, particularly since we also ask respondents to make decisions using the preferences of their spouse. Indeed we find that male respondents have a systematically biased view of their wives' labour supply preferences, overestimating their disutility of doing paid work. We incorporate this bias in our model and correct for it when explaining joint retirement. We then validate our SP data and the structural model by comparing model predictions with expected retirement ages directly reported by the same individuals. We find that the SP data have substantial predictive power: individuals who prefer later retirement according to the stated preference model, also tend to report a later expected age of retirement.

Our estimates of the structural model give plausible results. In line with the literature, the neo-classical unitary model is rejected and a higher relative wage rate increase bargaining power within the couple. If wage rates of husband and wife are equal, we do not find evidence of differences in bargaining power. We find significant evidence of (positive) complementarities in both partners' leisure and we find very strong evidence of a positive correlation between

partners' labour supply preferences. Our simulations suggest that correlation between preferences explain a much larger part of joint retirement than complementarities in leisure. On the other hand, the percentage of couples retiring jointly also strongly depends on budget opportunities (pension levels and accruals), even if there is no explicit link between an individual's pension and the partner's labour supply or retirement decision. Our simulation results should therefore interpreted as a qualitative illustration only, the magnitudes of the effects will be different if we use people's actual choice sets instead of the stylized scenarios in our current analysis. Yet, this papers shows that the collective model is perfectly suited to allow for a realistic representation of household behavior in a context with subjective expectations and imperfect knowledge of preferences.

References

- Banks, J., Blundell, R., and Casanova Rivas, M. (2007). The dynamics of retirement behavior in couples: Reduced-form evidence from England and the US. Mimeo, University College London.
- Barsky, R. B., Kimball, M. S., Juster, F. T., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement survey. *Quarterly Journal of Economics*, 112:537 – 579.
- Blau, D. M. (1998). Labor force dynamics of older married couples. *Journal of Labor Economics*, 16(3):595–629.
- Bourguignon, F., Browning, M., Chiappori, P.-A., and Lechene, V. (1993). Intra household allocation of consumption: A model and some evidence from French data. *Annales d'Economie et de Statistique*, 29:137–156.

- Casanova Rivas, M. (2010). Happy together: A structural model of couples joint retirement choices. Unpublished Manuscript, Department of Economics, UCLA, Los Angeles.
- Cherchye, L., De Rock, B., and Vermeulen, F. (2009). Opening the black box of intra-household decision making: Theory and nonparametric empirical tests of general collective consumption models. *Journal of Political Economy*, 117(6):1074–1104.
- Chiappori, P.-A. (1988). Rational household labor supply. *Econometrica*, 56(1):63–90.
- Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, 100(3):437–467.
- Chiappori, P.-A., Fortin, B., and Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, 110(1):37–72.
- Coile, C. (2004). Retirement incentives and couples’ retirement decisions. *Topics in Economic Analysis & Policy*, 4(1):Article 17.
- Couper, M. P., Ofstedal, M. B., and Lee, S. (2013). Encouraging record use for financial asset questions in a web survey. *Journal of Survey Statistics and Methodology*, 1(2):171–182.
- Delavande, A. and Rohwedder, S. (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly*, 72(5):866 – 891.
- Elsayed, A., de Grip, A., Fouarge, D., and Montizaan, R. (2018). Gradual retirement, financial incentives, and labour supply of older workers: Evidence from a stated preference analysis. *Journal of Economic Behavior and Organization*, 150:277–294.
- Fortin, B. and Lacroix, G. (1997). A test of the unitary and collective models of household labour supply. *The Economic Journal*, 107(443):933–955.

- French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies*, 72(2):395–427.
- Gustman, A. L. and Steinmeier, T. (2009). Integrating retirement models. NBER Working Paper No. 15607, National Bureau of Economic Research.
- Gustman, A. L. and Steinmeier, T. L. (2000). Retirement in dual-career families: a structural model. *Journal of Labor Economics*, 18(3):503–545.
- Gustman, A. L. and Steinmeier, T. L. (2004). Social security, pensions and retirement behaviour within the family. *Journal of Applied Econometrics*, 19(6):723–737.
- Hospido, L. and Zamarro, G. (2014). Retirement patterns of couples in Europe. *IZA Journal of European Labor Studies*, 3(12):1–18.
- Hurd, M. D. (1989). The joint retirement decision of husbands and wives. *Social Security Bulletin*, 54(1):29–32.
- Hurd, M. D. (2009). Subjective probabilities in household surveys. *Annual Review of Economics*, 1(1):543–562.
- Kapteyn, A. and Kooreman, P. (1992). Household labor supply: What kind of data can tell us how many decision makers there are? *European Economic Review*, 36(2):365–371.
- Kooreman, P. and Kapteyn, A. (1990). On the empirical implementation of some game theoretic models of household labor supply. *Journal of Human Resources*, 25(4):584–598.
- Levy, H., Janke, A. T., and Langa, K. M. (2015). Health literacy and the digital divide among older Americans. *Journal of General Internal Medicine*, 30(3):284–289.
- Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press.

- Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.
- McCarty, T. A. (1990). The effect of social security on married women’s labor force participation. *National Tax Journal*, 43(1):95–110.
- Michaud, P.-C. and Vermeulen, F. (2011). A collective labor supply model with complementarities in leisure: Identification and estimation by means of panel data. *Labour Economics*, 18(2):159–167.
- Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households’ choices of appliance efficiency level. *The Review of Economics and Statistics*, 80(4):647–657.
- Smith, J. P. (2011). Reconstructing childhood health histories. *Demography*, 46(2):387–403.
- Stancanelli, E. G. F. and van Soest, A. (2016). Partners’ Leisure Time Truly Together upon Retirement. *IZA Journal of Labor Policy*, 5(12).
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Van Soest, A. and Vonkova, H. (2014). How sensitive are retirement decisions to financial incentives? a stated preference analysis. *Journal of Applied Econometrics*, 29(2):246–264.

Figures

When you are between 62 and 65	When you are between 65 and 68	When you are 68 or older
You work 40 hours per week	You work 24 hours per week; your net income is the same as before age 65	You are retired and get a pension corresponding to 75% of your net earnings before age 65
Your partner is between 60 and 63	Your partner is between 63 and 66	Your partner is 66 or older
Your partner works 40 hours per week	Your partner is retired and gets a pension corresponding to 75% of [his/her] earnings before age 63	Your partner is retired and gets a pension corresponding to 75% of [his/her] earnings before age 63

Figure 1: Screen Shot of Stated-preference Scenario

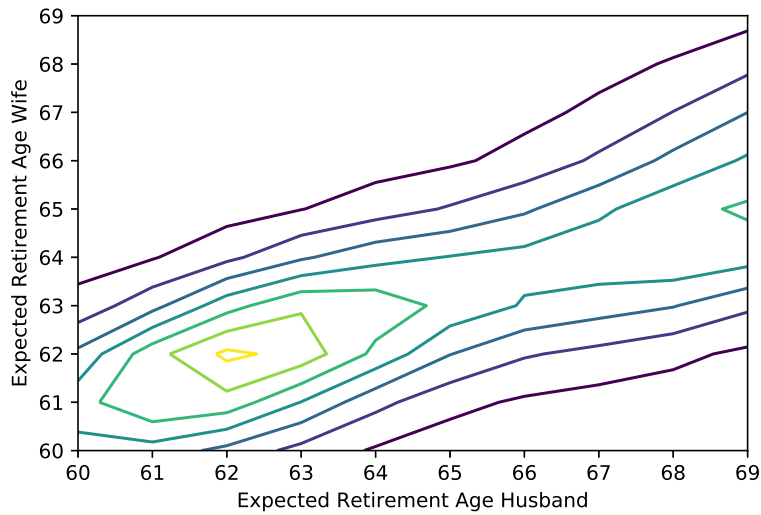


Figure 2: Simulated Distribution of Retirement Ages, Baseline Model

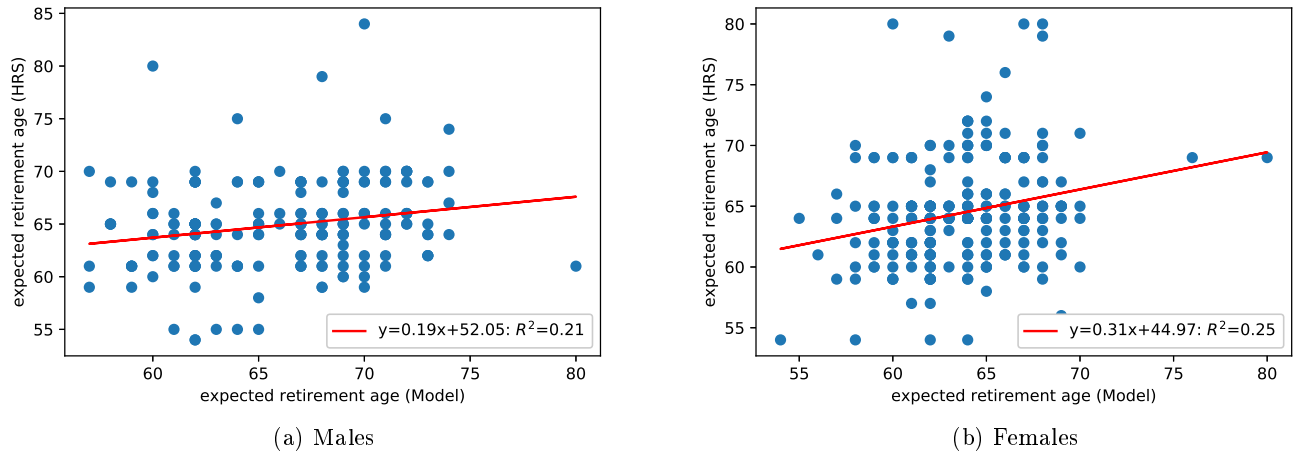


Figure 3: **Simulated Retirement Ages and Expected Retirement Ages in HRS**

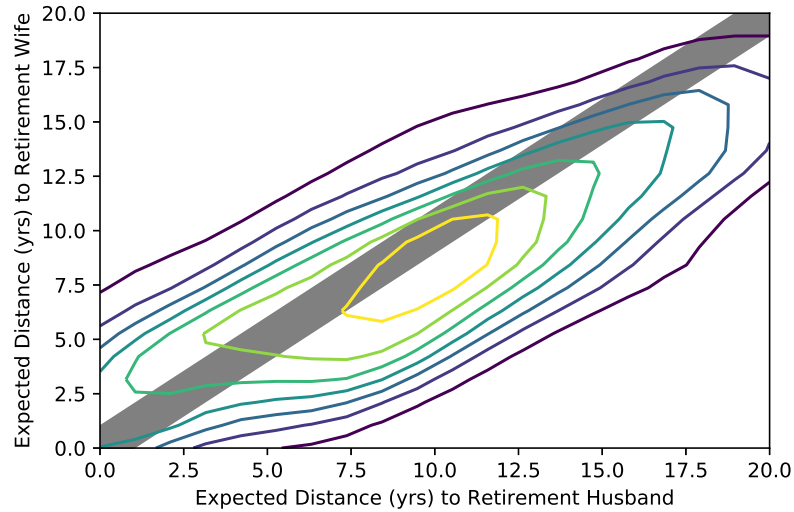


Figure 4: **Distribution of Simulated Distance to Retirement Husbands and Wives:** Shaded area represents distances that are at most one year apart.

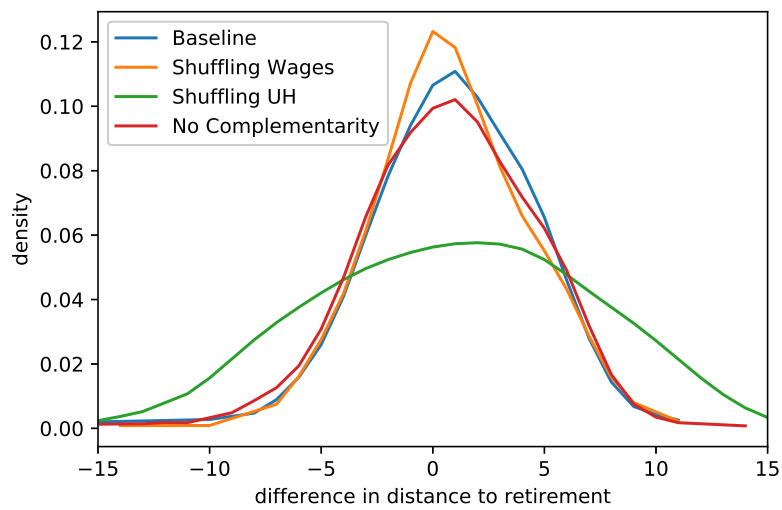


Figure 5: Simulated Distributions of Difference between Timing of Retirement of Husband and Wife.

Tables

	wives		husbands	
	mean	sd	mean	sd
age	54.54	4.49	56.38	4.18
any health limitations	0.09	0.28	0.07	0.25
college educated	0.42	0.49	0.44	0.50
job satisfaction	0.75	0.43	0.75	0.43
hourly wage	25.22	53.69	31.07	31.81
prov lives to 75 (relative to life-table)	0.93	0.25	0.97	0.32
probability has health limitations at age 62	0.39	0.26	0.40	0.27
Observations	604		604	

Table 1: **Summary Statistics Explanatory Variables:** Missing wages are imputed using OLS predictions from previous wages, age and education. Job satisfaction determined using whether respondent is satisfied or very satisfied with job. Subjective survival probabilities to age 75 reported as a ratio of the life-table probability by age and sex.

Scenario	Retirement Age						
	Male	Female	G1	G2	G3	G4	G5
1	65	68	50 60	55 65	60 70	65 75	70 80
2	62	65	40 40	45 45	50 50	55 55	60 60
3	68	68	60 50	65 55	70 60	75 65	80 70
4	62 (P) 65	65	50 40	55 45	60 50	65 55	70 60
5	68	65 (P) 68	70 60	75 65	80 70	85 75	90 80
6	65	65	50 50	55 55	60 60	65 65	70 70

Table 2: **Description of Retirement Scenarios in SP Questions by Randomized Group using an Example:** In Scenario 4 and 5, P refers to partial retirement. If the respondents or the spouses partially retire, their disposable incomes do not change. Their working hours reduce by 40 percent. Respondents were randomly assigned to one of the five groups G_1, G_2, G_3, G_4, G_5 , with different replacement rates in all scenarios. The first element in each group column gives the RR of the respondent while the second gives the RR of the spouse.

	Female respondents			Male respondents		
Made for	Scenario 2	Scenario 4	Skipped	Scenario 2	Scenario 4	Skipped
Males	0.324	0.568	0.108	0.296	0.632	0.071
Females	0.367	0.512	0.120	0.357	0.564	0.079
Both	0.349	0.531	0.120	0.296	0.629	0.075
	Scenario 5	Scenario 2	Skipped	Scenario 5	Scenario 2	Skipped
Males	0.454	0.420	0.127	0.518	0.389	0.093
Females	0.469	0.401	0.130	0.471	0.436	0.093
Both	0.478	0.392	0.130	0.511	0.400	0.089
	Scenario 3	Scenario 6	Skipped	Scenario 3	Scenario 6	Skipped
Males	0.293	0.571	0.136	0.425	0.493	0.082
Females	0.312	0.534	0.154	0.436	0.479	0.086
Both	0.306	0.549	0.145	0.443	0.464	0.093
Observations	324			280		

Table 3: **Proportion of Choices Made by Males and Females:** Fraction of respondents (column) who have chosen a particular scenario (three choice situations) for either one of the spouses, or both (rows). In the first choice situation, respondents chose between scenarios 2 and 4 (they could skip), the second, between scenarios 5 and 2 and finally in the third, between scenarios 3 and 6.

	Males	Females
Own leisure	α_i^m	β_i^f
constant	-1.686 (0.476)	-0.738 (0.353)
age	0.434 (0.065)	0.472 (0.085)
health limitations	1.609 (0.581)	1.411 (0.502)
college	0.168 (0.294)	-0.198 (0.238)
spouse responded	0.188 (0.350)	0.986 (0.357)
satisfied job	0.095 (0.348)	-0.209 (0.258)
leisure spouse	0.134 (0.023)	0.080 (0.022)
Consumption	α^c	β^c
	0.154 (0.037)	0.115 (0.037)
Discount factors	ρ^m	ρ^f
	1.030 (0.010)	1.018 (0.013)
Weights	λ	
Constant	0.014	
	0.330	
$\log(\frac{w_m}{w_f})$	0.517	
	0.308	
Heterogeneity	η_i^m	ν_i^f
Std.Dev	2.801	1.931
Correlation	0.995	
$\log L$	-2818.3	

Table 4: **Parameter Estimates:** Maximum likelihood estimates and standard errors for the collective model with subjective survival expectations.

	Loglikelihood Value	LR Statistic
Baseline	-2818.345	0.000
Fixed Discount Rates (2)	-2846.720	56.749
No Correlation UH (1)	-2899.353	162.016
No Complementarity (2)	-2855.766	74.843
Unitary (1)	-2821.128	5.567

Table 5: **Loglikelihood Ratio Test Statistics for Restricted Models:** number of restrictions in parenthesis. The critical value at the 5% level for 1 restriction is 3.84 and 5.99 for two restrictions.

	Ret Age Males	Ret Age Females	Fraction Joint
Baseline	65.728	63.830	0.336
Shuffling Wages	65.629	63.804	0.369
Shuffling UH	65.765	63.892	0.177
No Complementarity	66.142	64.472	0.304

Table 6: **Simulated Retirement Ages and Fraction Retiring at Most One Year from Each Other: Explaining Joint Retirement**

	Ret Age Males	Ret Age Females	Fraction Retiring Jointly
Baseline	65.728	63.830	0.336
No Penalty	63.759	62.244	0.435
High Penalty	66.478	64.399	0.276
Low Generosity	66.155	64.095	0.267
High Generosity	65.453	63.603	0.362

Table 7: **Simulated Retirement Ages and Fraction Retiring at Most One Year from Each Other: Financial Incentives:** The High Penalty scenario refers to an Actuarial Reduction Factor of 9% and Delayed Retirement Credit of 11% (compared to 6% and 8% in the baseline). The low generosity scenario refers to a replacement rate of 40% while the high generosity scenario refers to a replacement rate of 80%

8 Appendix: Estimation Results Alternative Specifications

	Males	Females
Own leisure	α_i^m	β_i^f
constant	-3.001 (0.904)	-1.027 (0.691)
age	0.697 (0.111)	0.907 (0.126)
health limitations	2.946 (1.023)	2.528 (0.945)
college	0.359 (0.555)	-0.443 (0.467)
spouse responded	0.440 (0.687)	1.753 (0.641)
satisfied job	0.208 (0.689)	-0.612 (0.511)
leisure spouse	0.310 (0.038)	0.168 (0.033)
Consumption	α^c 0.666 (0.056)	β^c 0.429 (0.042)
Discount factors	ρ^m 0.950 (0.000)	ρ^f 0.950 (0.000)
Weights	λ	
Constant	0.006 0.328	
$\log(\frac{w_m}{w_f})$	0.521 0.294	
Heterogeneity	η_i^m	ν_i^f
Std.Dev	5.179	3.763
Correlation	0.999	
$\log L$	-2846.7	

Table 8: **Parameter Estimates with Fixed Discount Factors:** Maximum likelihood estimates and standard errors for the collective model with subjective survival expectations.

	Males	Females
Own leisure	α_i^m	β_i^f
constant	-1.534 (0.479)	-0.842 (0.378)
age	0.400 (0.059)	0.445 (0.081)
health limitations	1.480 (0.607)	1.268 (0.495)
college	0.219 (0.299)	-0.076 (0.228)
spouse responded	0.223 (0.352)	0.807 (0.339)
satisfied job	0.129 (0.342)	-0.220 (0.262)
leisure spouse	0.119 (0.021)	0.075 (0.021)
Consumption	α^c 0.127 (0.031)	β^c 0.094 (0.031)
Discount factors	ρ^m 1.039 (0.010)	ρ^f 1.026 (0.012)
Weights	λ	
Constant	0.092 0.335	
$\log(\frac{w_m}{w_f})$	0.391 0.291	
Heterogeneity	η_i^m	ν_i^f
Std.Dev	2.689	1.760
Correlation	0.000	
$\log L$	-2899.4	

Table 9: **Parameter Estimates without Correlation in Unobserved Heterogeneity:** Maximum likelihood estimates and standard errors for the collective model.

	Males	Females
Own leisure	α_i^m	β_i^f
constant	-1.217 (0.401)	-0.433 (0.297)
age	0.455 (0.068)	0.457 (0.081)
health limitations	1.360 (0.523)	1.242 (0.447)
college	0.094 (0.257)	-0.190 (0.211)
spouse responded	0.543 (0.304)	1.170 (0.323)
satisfied job	0.123 (0.305)	-0.152 (0.226)
leisure spouse	0.000 (0.000)	0.000 (0.000)
Consumption	α^c 0.068 (0.021)	β^c 0.064 (0.021)
Discount factors	ρ^m 1.045 (0.012)	ρ^f 1.030 (0.013)
Weights	λ	
Constant	-0.064 0.320	
$\log(\frac{w_m}{w_f})$	0.533 0.290	
Heterogeneity	η_i^m	ν_i^f
Std.Dev	2.462	1.719
Correlation	0.996	
$\log L$	-2855.8	

Table 10: **Parameter Estimates without Complementarity in Leisure:** Maximum likelihood estimates and standard errors for the collective model.

	Males	Females
Own leisure	α_i^m	β_i^f
constant	-1.633 (0.479)	-0.761 (0.354)
age	0.440 (0.066)	0.476 (0.087)
health limitations	1.602 (0.580)	1.400 (0.502)
college	0.189 (0.295)	-0.190 (0.240)
spouse responded	0.172 (0.353)	0.955 (0.357)
satisfied job	0.031 (0.351)	-0.200 (0.260)
leisure spouse	0.132 (0.022)	0.083 (0.022)
Consumption	α^c	β^c
	0.152 (0.037)	0.116 (0.037)
Discount factors	ρ^m	ρ^f
	1.031 (0.010)	1.018 (0.013)
Weights	λ	
Constant	0.054	
	0.314	
$\log(\frac{w_m}{w_f})$	0.000	
	0.000	
Heterogeneity	η_i^m	ν_i^f
Std.Dev	2.817	1.944
Correlation	0.995	
$\log L$	-2821.1	

Table 11: **Parameter Estimates of Unitary Model:** Maximum likelihood estimates and standard errors