Labour Economics 85 (2023) 102431

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Retirement coordination and leisure complementarity

Irina Merkurieva[1]

School of Economics and Finance, University of St Andrews, Castlecliffe, The Scores, St Andrews, KY169AR, UK

ARTICLE INFO

JEL classification:
J26
J22
D13
J14

Keywords:
Retirement
Leisure complementarity
Intertemporal household choice

ABSTRACT

Empirical evidence suggests that in a large fraction of working couples spouses retire within a short period of time. This retirement coordination is frequently attributed to leisure complementarities. Contrary to this view, I find strong substitutability between the leisure of the two household members. Using a dynamic programming model of optimal retirement and labor supply decisions, I further show that high levels of retirement coordination can be observed even in the absence of leisure complementarity. Retirement coordination is higher in the households with more equality in the earnings profiles and utility derived from the leisure of their individual members.

1. Introduction

The problem of retirement from the labor force is increasingly important in the context of population aging and growing pressure on the old-age support systems. Historically retirement literature mainly focused on the individual lifetime labor supply problem. However as currently the largest group of older households is represented by dual-career married couples, retirement is more broadly recognized and modeled as an outcome of the joint household decision making process.

One of the most robust empirical facts describing the household retirement behavior is that couples seem to coordinate their time of retirement from the labor force. Some of the earliest papers to document this fact are Blau (1998) and Gustman and Steinmeier (2000). Based on the Health and Retirement Study (HRS) data, up to one third of older married couples in the U.S. leave the labor force in the same year. Similar patterns of retirement coordination have been documented across majority of developed countries.[2]

The most popular explanation of retirement coordination in the literature is complementarity between the leisure of husband and wife (Kniesner, 1976) or the “loving couple” assumption (Browning et al., 1985). In the presence of leisure complementarities, couples tend to retire simultaneously because of the utility gains from spending time together. Recent papers that find the evidence of preference for shared leisure or complementarities in leisure times of older married couples include An et al. (2004), Atalay et al. (2019), Bingley and Lanot (2007), Coile (2004), Michaud et al. (2020), Michaud and Vermeulen (2011), Schirle (2008), van der Klaauw and Wolpin (2008). Numerous further papers name leisure complementarity among likely explanations of retirement coordination, although do not necessarily aim to establish its presence (e.g., Kapur and Rogowski, 2007; Lalive and Parrotta, 2017; Stancanelli, 2017).

Methods used to identify leisure complementarity in these papers and the resulting estimates vary substantially. In many instances leisure complementarity is inferred implicitly either from an individual response to financial and policy incentives to the retirement of a spouse or from the interdependent individual labor supply decisions. In this paper I propose and implement a new test of leisure complementarity that is directly based on the estimated elasticity of substitution between individual leisure times. The test is derived from a life cycle model of labor supply with nested constant elasticity of substitution (CES) utility in consumption and leisure. I exploit the first-order conditions of the household problem to estimate the elasticity of substitution between the leisure of household members using the HRS data on individual labor supply and wages.

Contrary to the leisure complementarity hypothesis, my estimates suggest fairly high degree of substitutability between the leisure of husband and wife. The estimates of preferred model specification place the values of the elasticity of substitution in the range between 1.1 and 3.4. The test overwhelmingly rejects the null hypothesis of leisure complementarity and thus implies that retirement coordination must originate elsewhere. I further use the same preferred estimates of the

E-mail address: i.merkurieva@st-andrews.ac.uk.

[1] I thank John Kennan, Rasmus Lentz, Christopher Taber and James Walker for their helpful comments. All remaining errors are my own.

[2] Examples of papers that document retirement coordination around the world include (Banks et al., 2010) for the US and England, Schirle (2008) for Canada, Mastrogiacomo et al. (2004) for the Netherlands, Gustafson (2017) for Sweden, Hospido and Zamarro (2014) for the European countries participating in the Survey of Health, Aging and Retirement in Europe (SHARE) and Warren (2015) for Australia.

https://doi.org/10.1016/j.labeco.2023.102431
Received 2 March 2022; Received in revised form 5 July 2023; Accepted 7 August 2023
Available online 22 August 2023

household utility function to show that leisure complementarity is not required for the underlying model to generate coordinated retirement outcomes that are consistent with the data.

The results stem from a dynamic programming model of optimal labor supply and retirement behavior of older married couples. The model is calibrated to target the HRS data moments that describe individual labor supply decisions. The levels of retirement coordination generated by the model are consistent with the data. Coordination is linked to three distinct channels: income shocks, Social Security policy and the distribution of preferences for leisure across households. Using a set of counterfactuals, I show that while all suggested channels affect the level of retirement coordination, it is primarily linked to the relative contributions of individual leisure times to the household utility and the differences in the earnings profiles of the household members. Couples with less pronounced gender differences would exhibit more coordinated retirement behavior. In particular, closing the gap in the lifetime earnings would increase observed retirement coordination by one third.

In terms of the generated retirement coordination the model is observationally equivalent to the case of leisure complementarity, yet it offers a very different view of the household decision making. In a model with leisure complementarity, aging households would tend to decrease working hours of their members in proportion with individual contributions to the household utility. When leisure times are substitutable, households would start retirement by withdrawing the household member with lower earning capacity. The ability to distinguish between different sources of retirement coordination is essential in order to understand how policymaking will affect retirement from the labor force. Importantly, in both cases spousal retirement incentives as well as retirement state would be positively related to the likelihood of own retirement. Therefore, inferring leisure complementarity from the relationships between individual labor supply decisions may be misleading.

Although the outcome of leisure complementarity test in this paper opposes the most conventional explanation of retirement coordination, it is consistent with recent developments in the theoretical and empirical literature on household labor supply and time allocation. Estimates of the elasticity of substitution between the leisure of household members in this paper are of the same magnitude as those reported by Rogerson and Wallenius (2019) who also find this elasticity to be quite large and suggest that older households are willing to substitute leisure across their members. Knowles (2013) obtains similarly high elasticities of substitution in home production and labor of husband and wife for working age couples. The result that couples are less inclined to value shared leisure than previously thought increasingly emerges in the survey data. Eismann et al. (2017) report that majority of dual career couples name correlated preferences for leisure rather than preferences for shared leisure as their motivation to coordinate retirement. Based on the time use data, Stancanelli and Van Soest (2016) conclude that the actual amount of time couples spend together increases only insignificantly following joint retirement. I contribute to this growing literature by using a structural model of household retirement that helps understand how coordination works in the absence of leisure complementarity.

The rest of the paper contains six sections. The next section documents key facts about retirement coordination in the data. Section three explains the model. I propose and implement an empirical test of leisure complementarity in Section 4. In Section 5 I calibrate a structural model of the two-member household labor supply and retirement decision, explore the channels of retirement coordination and describe model predictions and counterfactuals. Section 6 concludes the paper.

2. The evidence of retirement coordination

The first papers to document the prevalence of joint retirement are based on the data that roughly correspond to the period between 1960 and 1990 (Blau, 1998; Gustman and Steinmeier, 2000; Hurd, 1990). Their findings cannot be taken for granted in more recent periods because of socially driven developments in the labor market behavior of older household, most important the sustained growth in female attachment to the labor force over the life cycle. A comparison of employment rate among females in the 55–59 age group who were on the edge of retirement in 1985 to that of the next generation 25 years later reveals an increase in excess of one third.[3] This is a dramatic change that inevitably affected the household retirement decision making and possibly changed the extent of observed retirement coordination. Therefore, in this section I start the data analysis by verifying that, in spite of the major shifts in female employment, retirement coordination is still present in contemporary data.

Throughout the paper I use the data from nine core waves of the Health and Retirement Study (HRS) conducted in 2000–2016 and the corresponding off-year Consumption and Activities Mail Surveys (CAMS) from 2001–2017.[4] The HRS is a nationally representative longitudinal study of the U.S. population over the age of 50. The data are collected biennially since 1992 and cover a broad range of subjects including employment, earnings and wealth, family structure, participation in the government programs, health and mortality. The time range in this paper is constrained by the availability of consumption data that are required for complete estimation of the household preferences.

The estimation sample includes heterosexual married and cohabiting couples interviewed in the community. In addition to the missing data, the sample is further restricted to households in which age difference between the spouses does not exceed fifteen years. Each household member in the sample is required to have at least five years of job market experience over the lifetime, removing single career households where retirement is virtually an individual rather than joint decision. The resulting sample is an unbalanced panel of 34,421 household-year observations for 7,994 unique households. Table 1 shows descriptive statistics for the main variables used in the paper.

The idea of joint retirement seems to resonate with the HRS households. The first respondents to be surveyed back in 1992, at the time on average around 55 years old and predominantly not yet retired, were asked about their retirement plans. The preference for joint retirement was expressed by 70.6% of non-retired households where at least one person claimed their intention to retire together with the spouse. As these households are followed up over a number of years, we get to observe the actual retirement outcomes for approximately two thirds of the original sample. Although their high expectations of joint retirement are not quite fulfilled, there is still a fair level of coordination in the timing of retirement. Judging by self-reported retirement status, partners in 34% of retired households announced their first-time retirements in the same survey waves.

Although subjective definition of retirement by self-report is consistent with the data on individual expectations, it is problematic for use in the formal modeling setting. To avoid possible issues around the ambiguity of subjective definition and its openness to individual interpretation, for the purposes of this paper I define retirement using the labor force status. Anyone out of the labor force at the time of their interview is considered retired, with the time of retirement determined by the end date of the last recorded job. For multiple transitions in and out of the labor force, the retirement date is determined by the earliest exit registered in the data.[5]

[3] Computed using the Labor Force Statistics from the Current Population Survey (CPS).

[4] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant numberNIA U01AG009740) and is conducted by the University of Michigan.

[5] Descriptive results for retirement coordination are not sensitive to the choice between different exit dates, e.g. earliest or latest observed exit from the labor force.

Table 1
Descriptive statistics.

Variable	Males			Females		
	Mean	Median	S.D.	Mean	Median	S.D.
Age, years	65.5	64.0	9.0	64.3	63.0	8.4
Black, %	6.3			5.6		
Hispanic, %	5.7			5.7		
Schooling, years	13.5	14.0	2.9	13.4	13.0	2.5
Good health, %	78.5			81.1		
No chronic health conditions, %	17.7			18.1		
Employed, %	51.9			44.7		
Employed full-time, %	39.2			27.0		
Not in the labor force, %	46.1			53.6		
Age of the first labor force exit	64.2	63.1	6.4	63.4	62.9	7.1
Age of self-reported retirement	61.2	62.0	6.3	61.8	62.2	7.3
Observed retirement outcome, %	52.2			51.6		
Annual hours of work	2,014	2,080	831	1,681	1,920	772
Hourly wage	25.3	18.2	26.9	17.4	12.8	17.5
Average annual earnings	55,880	42,078	57,743	32,704	25,699	27,450
Years worked	40.3	41.0	10.4	30.8	32.0	11.9
Receive Social Security, %	66.8			54.7		
Age at Social Security take up	62.3	62.3	4.8	62.0	62.2	4.7
Social Security income	14,542	14,400	6,534	9,401	8,400	5,344

	Household		
	Mean	Median	S.D.
Age difference in months, male–female	31.7	28	54.8
Total income	73,700	54,752	64,846
Value of housing and financial wealth	453,300	242,500	602,033
Consumption	49,970	43,128	30,300
Number of household residents	2.5	2.0	1.0
Number of financial dependents	0.27	0	0.76
Resident child, %	27.7		
Have grandchildren, %	75.0		
Number of periods in the sample	4.4	4.0	2.4
Number of household observations		34,421	
Number of unique couples in the sample		7,994	

Notes: Pooled statistics for 2000–2016 estimation sample: coupled households with at least five years of job market experience and age difference under 15 years; observations with missing data are excluded. Results are weighted by the HRS individual and household weights. All monetary values are converted to 2000 US dollars. Income, wages and hours statistics are conditional on participation. Health indicator is a dummy variable determined by self-report (1 = excellent, very good or good).

Based on the labor force status definition of retirement, the households where two spouses quit work within one year of each other represent 27% of retirements in the follow up of the original sample; almost a third of these joint retirements are dated by the same month. These proportions are similar in the estimation sample that includes households from subsequently added cohorts. Considering that retirement is only a part of the household labor supply problem, it is useful to look for the signs of general labor supply coordination. It turns out that spouses share the same retirement status in 71% of available observations, comprised of 33% working and 38% retired households. Further facts that point at the coordination of labor supply are the differences in the partners' weekly hours of work that are under five hours in 48% of the couple-year observations and the substantial fraction of retired couples (33.5%) who took up Social Security in the same calendar year. Overall, it seems that up to one third of the HRS couples chose to retire together.

Even though these statistics suggest that joint retirement and synchronization of labor supply decisions are commonplace in the data, it is possible that observed coordination merely reflects the distribution of age differences in the households. After all, we are looking at a sample of older workers, and it would not be too surprising to learn that people of roughly similar age coincidentally retire around the same time. To explore this possibility, Fig. 1 plots the distribution of differences between retirement dates of the spouses along with the distributions of their age differences and the time between the dates of Social Security take up. The distribution of differences between retirement dates is almost symmetric around zero. A clear peak at smaller distances between the months of labor force exit confirms the presence of joint retirement in the data. The other two distributions are visibly skewed in the opposite directions: age differences to the right and the differences in Social Security take up dates to the left. In both cases, the skewness is due to the fact that male partners in the couples are on average slightly older. Importantly, unlike the Social Security take up that is largely driven by age due to policy incentives, the differences in retirement dates do not quite follow the age profile of older households.

To formalize the intuition behind this relationship, I estimate a reduced form linear probability model of the individual retirement decision as a function of retirement status of the partner, couple's age difference and personal characteristics (age quadratic, education, work experience and health). Table 2 shows several sets of estimates for this relationship. The main estimates in columns (1)-(2) are the maximum likelihood estimates of two simultaneous retirement equations in which retirement status of the two partners is determined jointly. For comparison, columns (3)-(4) contain the estimates of baseline equations for each gender with independent retirement decisions. Models in columns (5)-(6) estimate the likelihood of retirement treating the labor force status of a spouse as exogenous.

The main conclusion from these estimates is that the partner's retirement status is an important and statistically significant predictor of the retirement probability even after controlling for age differences and other variables in the model. At the age of 65, the effect of partner's retirement in the simultaneous equations model is equivalent to 1.7 additional years of age for males and 4.2 years for females. That is, a 65-year-old female with retired husband would be as likely to be retired herself as an otherwise equivalent 69-year-old with working partner. As expected, these estimates are lower than in the model with exogenously determined partner's retirement. The latter would be biased upwards

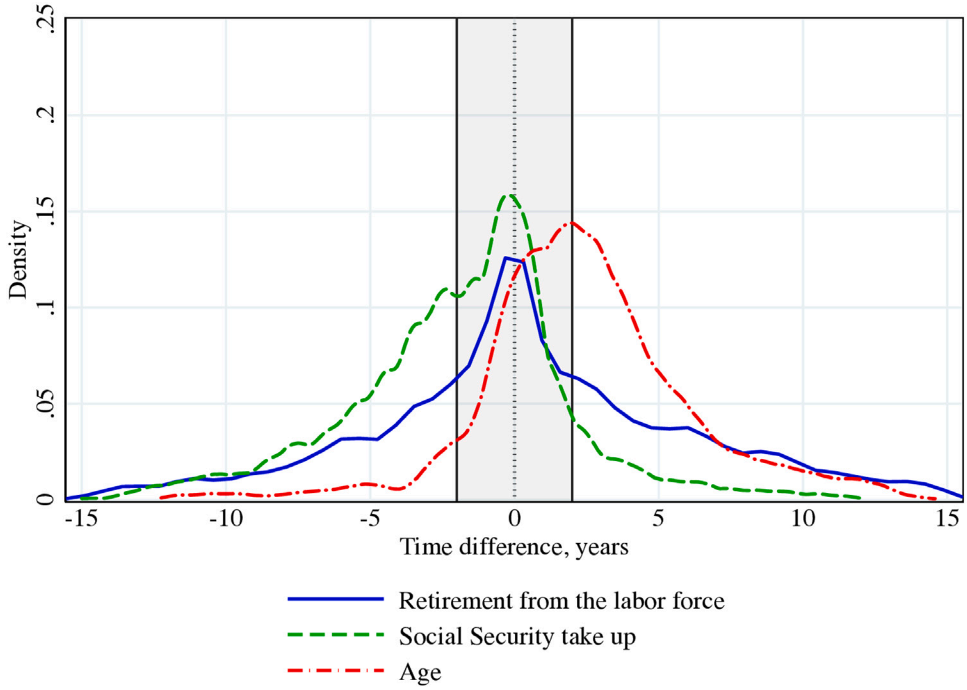

Fig. 1. The distribution of differences in the retirement dates of the spouses.
Notes: The plots show kernel density functions of the time differences in retirement, Social Security take up and age of the household members (Epanechnikov kernel, bandwidth $h = 0.5$). The time differences are measured as male minus female outcome in years; all differences are truncated at 15 years. Retirement status is defined by labor force participation; retirement date is the end date of the last job held prior to the first-time exit from the labor force. Estimation sample ($n = 1,347$): coupled households from the HRS 2000–2016 with at least five years of job market experience; observations with missing data and incomplete retirement histories are excluded. The vertical dotted line corresponds to zero time difference; the shaded area is the time frame of two years either side of zero.

due to endogeneity that arises from the joint nature of the household retirement decision. Even after accounting for simultaneity, the effect of partner's retirement is substantially larger than that of the age difference: for a 65-year-old worker one less year of age difference with the partner is equivalent to only two additional months of age. These results support the intuition derived from Fig. 1, and further suggest that the determinants of observed retirement coordination are more complicated than just the distribution of age differences within the households.

Much as it is difficult to draw direct comparisons to other papers because of methodological differences, these estimates of retirement coordination are not unlike those obtained for earlier periods. Blau (1998) computes that between 30.3% and 40.6% couples start retirement within one year of each other; my estimates tend to fall towards the lower end of this range. Similar to this paper, Gustman and Steinmeier (2000) find that for males the incentives from spouse's retirement are equivalent to approximately two years of age, albeit their findings for females are very different. Overall it seems that, in spite of all changes in social norms and labor market behavior, joint retirement is still extremely common in the contemporary data. Having documented this fact and shown that retirement coordination cannot be entirely attributed to the distribution of age differences, I now turn to the role of leisure complementarity in retirement coordination.

3. Model

The theoretical framework used to explore the origins of retirement coordination in this paper builds upon a dynamic life cycle model of household labor supply. The unit of analysis is a married household consisting of two members: husband and wife, indexed by $s = \{h, w\}$. Households are heterogeneous in their preferences over consumption and leisure, age differences between the spouses, membership in the private pension plans and lifetime earnings that account for the human capital of the household members. These variables collectively determine the household type. The key elements of the model for the household of given type are presented below.

3.1. Preferences and timing

A household may live for up to T periods starting from the moment its older member reaches the age of 50. In each period $t \in \{0, \dots, T\}$ it maximizes the joint utility from shared consumption, C_t, and individual leisure of the two household members, L_t^s. Preferences are additive over time and separable across the states of nature; consumption and leisure are normal goods.

The household's single-period utility function $U(\cdot)$ is assumed to take the form of the nested constant elasticity of substitution (CES) composite,

$$U(C_t, L_t) = \left[\alpha_t L_t^{\rho} + (1 - \alpha_t) C_t^{\rho}\right]^{1/\rho}, \tag{1}$$

where the term L_t aggregates utility derived from individual leisure allowances. It forms the inner nest of the utility function,

$$L_t = \left[\alpha_L (L_t^h)^{\rho_L} + (1 - \alpha_L)(L_t^w)^{\rho_L}\right]^{1/\rho_L}. \tag{2}$$

All four parameters of the utility function are type specific and reflect heterogeneity of preferences. In addition, the weight placed upon the household leisure aggregate, α_t, depends on time and varies with the age of individual household members. As households grow older, they value leisure more and eventually retire.

Before reaching the terminal age T, at which all individuals die with probability one, each household member faces an exogenous mortality risk. An individual of gender s alive at time t survives to period $t + 1$ with probability π_t^s. In case of the death of a household member, their

Table 2
Reduced form relationship between retirement decisions of the spouses.

Variable	Simultaneous retirement equations		Independent retirement equations			
			Independent retirement decisions		Interdependent retirement decisions	
	Males (1)	Females (2)	Males (3)	Females (4)	Males (5)	Females (6)
Spouse retired	0.063	0.116	–	–	0.167	0.171
	(0.020)	(0.019)			(0.009)	(0.009)
Age difference with the spouse, years	−0.005	−0.005	−0.006	−0.008	−0.002	−0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age, years	0.112	0.085	0.117	0.094	0.103	0.082
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age squared (×0.01)	−0.058	−0.045	−0.061	−0.049	−0.054	−0.043
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Education, years	−0.008	−0.002	−0.009	−0.004	−0.008	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience, years	−0.011	−0.010	−0.011	−0.010	−0.010	−0.010
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Good health	−0.139	−0.126	−0.140	−0.129	−0.138	−0.124
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Constant	−3.690	−2.702	−3.901	−3.019	−3.385	−2.570
	(0.161)	(0.132)	(0.147)	(0.123)	(0.144)	(0.122)
F-statistic	–		830	839	877	889
Log pseudolikelihood	−569,473		–	–	–	-

Notes: Dependent variable is a binary indicator of retirement defined by the labor force status (1 = individual is out of labor force). Models 1 and 2 report maximum likelihood estimates of two simultaneous retirement equations for male and female partners in the household. Models 3–6 report the least squares estimates of the linear probability models of individual retirement equations. Models 3–4 assume retirement decisions within the household are made independently; in models 5–6 individual retirement decision is allowed to depend on exogenously determined retirement status of the spouse. Health indicator is a dummy variable determined by self-report (1 = excellent, very good or good). Figures in the parentheses are robust standard errors clustered by household. All specifications include year fixed effects. Estimation sample includes coupled households from the HRS 2000–2016 with at least five years of job market experience and age difference under 15 years. Sample size $n = 34,348$.

partner inherits all accumulated assets and continues life as a single-member household. The preferences of single-member households are CES in consumption and leisure,

$$U^s(C_t, L_t^s) = \left[\alpha_t^s (L_t^s)^{\rho_s} + (1-\alpha_t^s) C_t^{\rho_s}\right]^{1/\rho_s}. \tag{3}$$

The decisions of single-member households are constrained by the individual mortality processes, as well as the budget constraints that are obtained from their household equivalents introduced below.

3.2. Budget constraints

A household receives income from three sources: employment, Social Security and retirement benefits, and returns on assets. Each household member can supply labor to the market out of a fixed endowment $\bar{L}$. Work is paid at individual wage rates W_t^s that are determined by exogenous random process. Retired household members allocate their entire time endowment to leisure. There are no frictions and unemployment, so retirement and not working are equivalent notions.

Upon reaching the statutory retirement age, individuals become eligible to receive the Social Security retirement and pension benefits S_t^s. The amount of benefits depends on the lifetime earnings of the household members and their age at take up. The timing of Social Security take up is determined exogenously so that to maximize the expected lifetime benefits collected by the household. The amount of benefit collected after take up is constant; future benefits are not recomputed. The latest take up age is 70, as the system offers no incentives for further delay. The average value of the lifetime earnings is a part of the individual type; it does not get updated with labor supply decisions after the age $t = 0$. It is possible to work and obtain Social Security retirement benefits at the same time. Pension benefits are modeled as a function of a worker's age and Social Security benefits. Further details of the Social Security policy and the discussion of assumptions made in the paper can be found in online Appendix A.

Households save and invest a joint stock of assets A_t at a constant interest rate r. There are no taxes in the model.[6] Households are not allowed to borrow against their future wage income or Social Security. The terminal condition sets $A_{T+1} = 0$; for all periods $t \le T$ household assets are determined by the budget constraints

$$A_{t+1} = (1+r)\left[A_t - C_t + \sum_{s=h,w}\left(\left(\bar{L} - L_t^s\right)W_t^s + S_t^s\right)\right]. \tag{4}$$

3.3. Recursive formulation

The vector of state variables at time t includes individual wage rates, Social Security status, household assets and survival state. Future wages and survival are uncertain. A household forms beliefs over the distribution of their values; these beliefs and discounting factor β are assumed to be identical across the household members. Agents have rational expectations and the state transition probabilities are conditionally independent. Conditional upon the random processes for individual wages and survival, the Social Security status and assets evolve deterministically.

Given the current state at time t and the household type, the household makes decisions on individual labor supply and joint consumption, $D_t = \{C_t, L_t^h, L_t^w\}$. The decision is chosen so that to maximize the expected discounted utility over the remaining lifetime subject to the exogenous processes for mortality and wages, budget constraints and Social Security rules.

Let the value of state $\Omega_t = \{W_t^h, W_t^w, S_t^h, S_t^w, A_t\}$ for a living married household be $V_t = V_t(\Omega_t, \theta)$, where θ is the vector of model parameters, including parameters of the state transition probability function. Similarly, define the state values for single member households of gender s as $V_t^s = V_t^s(\Omega_t, \theta)$. The household state values are defined recursively by

$$V_t = \max_{D_t}\Big\{ U_t(D_t) + \beta \int \left[\pi_t^h \pi_t^w V_{t+1} + (1-\pi_t^h)\pi_t^w V_{t+1}^w + (1-\pi_t^w)\pi_t^h V_{t+1}^h\right] \cdot p(d\Omega_{t+1}|\Omega_t, \theta)\Big\}, \tag{5}$$

where integration is over the distribution of future wage values. That is, the value of the state Ω_t comprises the current utility from consumption and leisure, plus the expected present value of the future state for the household that, depending on the realized survival process, may consist of either one or two members. Although this model does not have a closed form solution and has to be estimated numerically, it is possible to exploit its first order conditions in order to test the hypothesis of leisure complementarity.

4. The test of leisure complementarity

Selected functional form of the utility function does not require imposing ex ante assumptions on the substitutability between the leisure of the spouses. Parameter $\rho_L \in (-\infty, 1)$ in the inner nest of the CES utility (2) characterizes the elasticity of substitution between the leisure of husband and wife, σ_L, as $\sigma_L = \frac{1}{1-\rho_L}$. The limiting values of ρ_L yield the cases of perfect substitutability of leisure ($\rho_L = 1$), perfect complementarity ($\rho_L = -\infty$), and Cobb–Douglas preferences when relative demands for goods are independent of relative prices ($\rho_L = 0$). Beyond

[6] The main estimates are based on the ratio of household wages. Most couples have incentives to file joint tax returns, so that the average tax rate is the same for both household members. Because in such cases the estimates will not be affected by income tax, the assumption has little impact on the results.

Table 3
Estimates of the wage equations.

	Low education		High education	
	Males	Females	Males	Females
Work experience, years	0.052	0.058	0.093	0.088
	(0.011)	(0.009)	(0.009)	(0.008)
Square of work experience (×0.01)	−0.046	−0.074	−0.105	−0.118
	(0.013)	(0.013)	(0.011)	(0.012)
Tenure, years	0.009	0.005	0.011	0.010
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.508	0.554	0.776	0.600
	(0.102)	(0.079)	(0.088)	(0.082)
Persistence of AR(1) component, ρ_v	0.333	0.344	0.323	0.298
Variance of AR(1) component, σ_v	0.686	0.629	0.756	0.667
F-statistic	31.5	40.5	50.9	57.1
	(0.000)	(0.000)	(0.000)	(0.000)
Number of observations	3,220	3,450	5,394	4,941

Notes: Fixed effects estimates with an AR(1) disturbance, Eq. (8). Dependent variable is log of real hourly wage. Figures in the parentheses are standard errors. All models include region and individual fixed effects.

the limiting values, the leisure of husband is a gross complement to the leisure of wife for values $\rho_L < 0$ that correspond to $0 < \sigma_L < 1$. In this case, as the relative price of husband's leisure increases, the relative amount of labor supplied by husband would increase as well, but proportionately less than the rise of relative price. The opposite happens for values $0 < \rho_L < 1$. Because in this case $\sigma_L > 1$, an increase in the husband's relative labor supply is proportionately larger than an increase in relative price, and the leisures of the household members are gross substitutes. The following reduced form approach allows to estimate ρ_L and derive conclusions about leisure complementarity without solving the entire model.

4.1. Reduced form test

I exploit the algebra of the CES preferences to estimate parameters of the household utility function as in Heckman et al. (1998). The log ratio of the first order conditions for individual leisure choices yields

$$\log \frac{W_t^h}{W_t^w} = \log \frac{\alpha_L}{1-\alpha_L} + (\rho_L - 1) \log \frac{L_t^h}{L_t^w}. \tag{6}$$

Eq. (6) highlights the role of pecuniary and non-pecuniary motives in the labor supply decisions of older households that are related to individual wages and preferences over leisure of the household members. For a given value of the wage gap, a couple that places higher weight on the husband's leisure time would have lower leisure gap, and therefore exhibit higher degree of retirement coordination, regardless of the degree of substitutability between individual leisure terms. Similarly, couples with lower wage gap will exhibit higher levels of retirement coordination holding parameters of the utility function α_L and ρ_L fixed. These observations highlight the key intuition behind some of the main paper results, namely that in the absence of leisure complementarity the degree of retirement coordination is primarily related to the weights households place on the leisure of their members and the wage gap within the couple. This intuition is consistent with the data, where we observe positive reduced form relationship between the household wage and leisure gaps (see online Appendix B for additional data facts).

Given the household panel data that contain information on individual earnings and labor supply, parameters α_L and ρ_L can be consistently estimated by the equation

$$\log \frac{W_{it}^h}{W_{it}^w} = \beta_{10} + \beta_{11} \log \frac{L_{it}^h}{L_{it}^w} + \phi_{1i} + \varepsilon_{1it}, \tag{7}$$

where ϕ_{1i} is time-invariant fixed effect for couple i and ε_{1it} is an i.i.d. error term that captures unexplained variation in the spouse wage gap. The null hypothesis of complementarity between the leisure of the household members can be explicitly formulated in terms of the estimated parameter as

$$H_0 : \beta_{11} \le -1$$

and tested by a standard t-test. The null hypothesis is consistent with the household leisure terms being gross complements in the household utility function. Rejecting the null would imply insufficient evidence of leisure complementarity in the data. In the next section I discuss the empirical results of the proposed test using the HRS dataset.

Although estimation of the inner CES nest by Eq. (7) is the only step required for the test of leisure complementarity, all remaining parameters of the household utility function can be obtained as shown in online Appendix C. I use this approach to retrieve complete set of the utility function parameters and simulate the household retirement decision in Section 5 of the paper.

4.2. Empirical implementation

The test of leisure complementarity is based on the first order conditions for the household problem, therefore the estimation sample excludes all data points that correspond to the corner solutions. Excluded households are those with non-working retirees as well as households where at least one member reportedly allocated the entire time endowment to work. This reduces the estimation sample described in Section 2 to 8,251 observations. The test implementation further requires specification of empirical counterparts to the theoretical variables that were introduced in the model. These are defined as follows.

The wage gap is computed from the individual wages of the household members. I use predicted wage rates instead of the values reported in the HRS data in order to address possible concerns about spurious correlation and division bias. The logarithm of real hourly wage earned by the individual i of gender s at time t, $\log W_{it}^s$, is assumed to follow an error component model

$$\log W_{it}^s = x_{it}^s \beta + f_i + v_{it}^s, \tag{8}$$

where x_{it}^s is a row vector of regressors and f_i is the individual fixed effect. The error term v_{it}^s is described by a stationary AR(1) process with persistence parameter ρ_{v^s},

$$v_{it}^s = \rho_{v^s} v_{it-1}^s + \zeta_{it}^s. \tag{9}$$

The innovations ζ_{it}^s are independent over individuals and time periods and are identically distributed with mean zero and variance $\sigma_{v^s}^2$. The vector of regressors x_{it}^s includes a quadratic function in the years of labor market experience, tenure at the current job and a set of region dummies. Wage equation is estimated separately by gender and education groups: high school degree or less ("low") and some college or more ("high"). Estimation results are reported in Table 3.

The annual time endowment is set at 8,766 hours, the number of hours in a calendar year. The annual hours of work are computed as the product of self-reported number of weeks worked in a year and the usual number of hours in a working week. Individual leisure amount is approximated by the difference between the time endowment and the annual hours of work, and the ratio of the individual leisure terms gives the household leisure gap.

It is important to notice that these particular specifications of the main variables do not affect the key outcomes of the leisure complementarity test. While all choices were guided by considerations that made them in some ways superior to the alternatives, the results are robust to modifications. In particular, the outcomes of leisure complementarity test will be the same if the actual wages were used in place of predicted or if the data included all observations rather than only internal solutions to the household labor supply problem. Furthermore, mechanical variations to the assumed annual time endowment have no

Table 4
Leisure complementarity in the household utility function.

	Model 1	Model 2	Model 3	Model 4
	Fixed effects	IV, excluded instruments:		
		Household structure	Health	All instruments
Estimated parameters of Eq. (7):				
Slope, β_{11}	−0.134	−0.599	−0.636	−0.613
	(0.028)	(0.223)	(0.212)	(0.164)
Intercept, β_{10}	0.219	0.197	0.195	0.196
	(0.002)	(0.014)	(0.011)	(0.010)
Point estimates of the CES parameters:				
Substitution b/w leisure of husband and wife, ρ_L	0.866	0.401	0.364	0.387
Average weight on husband's leisure, α_L	0.552	0.547	0.546	0.547
The test of leisure complementarity:				
95% C.I. for ρ_L	(0.81,0.92)	(−0.04,0.84)	(−0.05,0.78)	(0.07,0.71)
95% C.I. for σ_L	(5.3,12.5)	(0.97,6.2)	(0.95,4.6)	(1.1,3.4)
p-value for $H_0 : \rho_L \leq 0$	0.000	0.036	0.043	0.009
Post-estimation tests for IV models:				
Hansen-Sargan test	–	1.65	7.07	8.79
(p-value)		(0.895)	(0.794)	(0.947)
Endogeneity test	–	5.35	6.695	10.506
(p-value)		(0.021)	(0.010)	(0.001)
Underidentification test	–	38.7	29.8	62.0
(p-value)		(0.000)	(0.003)	(0.000)
Weak identification test (10% test size, 5% critical value)	–	6.51	3.70	4.35
		(4.45)	(3.50)	(3.20)
Number of observations	8,251	8,251	8,251	8,251

Notes: The table contains parameter estimates of the inner nest of the CES utility function by Eq. (7). Model 1 is estimated by fixed effects. Models 2–4 are estimated by IV with different sets of instruments: household structure (Model 2 — the number of grandchildren and financial dependents), health (Model 3 — ADL limitations, CESD score and stroke diagnoses) and both household and health instruments (Model 4). All IV specifications include couple fixed effects and are estimated by limited information maximum likelihood (LIML). The standard errors are robust and clustered by household. The standard errors for Model 1 are computed by panel nonparametric bootstrap and take into account the wage gap estimation. Estimation sample includes working coupled households from the HRS 2000–2016 with at least five years of job market experience and age difference under 15 years.

effect on the test outcome.[7] I show in the end of this section that neither does more advanced treatment of the leisure variable that accounts for the role of home production in the household time allocation decision.

In addition to these variables, complete estimation of the household utility function described in online Appendix C requires data on the household consumption. Consumption measure used in this paper is based on the total household consumption variable from the Consumption and Activity Mail Supplement (CAMS), a regular supplement to the main HRS administered since 2001. Because CAMS is only sent out to a random subsample of the core HRS respondents, the sample with complete data is very small. Missing consumption values were imputed from a linear regression of CAMS log consumption on the variables from the core survey, including the age of the household members, their total assets, education, labor supply, income, and the number of household residents.

4.3. Baseline results

All estimation results are summarized in Table 4. The top two sections of the table report the raw estimates of Eq. (7) and their standard errors, followed by the point estimates of the corresponding parameters of the household utility function. The weight placed on the leisure of each individual in a household is derived from the intercept and a couple fixed effect that cannot be identified separately. Estimated leisure weights are therefore couple-specific. In order to facilitate interpretation, the table of regression results reports the average estimated leisure weights computed as

$$\bar{\hat{\alpha}}_L = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_{Li} = \frac{1}{n}\sum_{i=1}^{n}\frac{\exp(\hat{\beta}_{10}+\hat{\phi}_{1i})}{1+\exp(\hat{\beta}_{10}+\hat{\phi}_{1i})}.$$

The first column of the table contains results for the baseline model specification in which Eq. (7) is estimated by fixed effects. The average weight on husband's leisure has an estimate of 0.552, so that couples value the leisure of household members almost equally. The point estimate $\hat{\rho}_L = 0.866$ implies the elasticity of substitution between the leisure of husband and wife $\hat{\sigma}_L = 7.5$, which is much higher than the cutoff value of one required for complementarity.

The lower section of the table provides a formal summary of the leisure complementarity test. In the baseline model the test strongly rejects the null with a p-value below 0.0001. High degree of substitutability between the leisure terms is particularly salient from the inspection of the 95% confidence interval for the parameter $0.81 < \rho_L < 0.92$ that contains a tight range of values distinctly away from zero. The test therefore offers an overwhelming evidence against the presence of leisure complementarity in the household preferences. As this outcome appears to contradict at least some of the extant literature and challenges the way in which we tend to think about the household retirement problem, I now address two model specification issues that could lead to erroneous estimates of the baseline model specifications.

4.4. IV estimation

A reasonable concern about consistency of the baseline estimates is due to the potential correlation between unobserved shocks to the spouse wage gap and the household labor supply decisions that determine the leisure gap. Measurement error in the leisure gap can also bias the estimates. In this section I address these problems by instrumenting the household leisure gap. I employ two conceptually different sets of instruments, first separately and then jointly.

The first set of instruments relates to the composition of the household. It is known that caretaking responsibilities differ substantially by gender, and so it is likely that the presence of immediate and extended household members who require support in either care or financial terms would be informative of the household leisure gap. Such factors are less likely to affect the labor productivity and wages that individuals can command in the market. To capture the relevant household characteristics, I include information on the number of financial dependents and grandchildren of the couple.

The second set of instruments relies upon individual health conditions. I use three data pieces: limitations on the activities of daily living (ADL), mental health and exposure to stroke. In the data these characteristics have the strongest impact on individual leisure choices when compared to the other available measures of health. There can be some concern that deteriorating health would equally decrease market wages, yet this is mitigated by using predicted wages that are net of the direct health impact. The HRS includes questions about five activities of daily living: walking across room, dressing, bathing, eating and getting in or out of bed. I define the relevant variable as the number of ADLs in which an individual experiences at least some difficulties. Mental health is measured by CESD score on a scale from 0 to 8, with higher values reflecting more negative feelings reported by the respondent over the past week. I capture the impact of mental health by a dummy that takes a value of 1 for anyone with non zero CESD score.

The results of IV estimation are reported in columns 2–4 of Table 4. Column 2 contains results obtained with the household instruments, column 3 that with health instruments, and column 3 shows estimates when both groups of instruments are used together. In addition to the results reported for the baseline model, I now record statistics that help

[7] Online Appendix D summarizes the main estimates for several values of time endowment $\bar{L}$.

assess performance of the proposed instruments in the bottom section of the table. In all three specifications the tests support the proposition of leisure gap endogeneity, and rule in favor of IV estimation. Hansen-Sargan overidentification test fails to reject the null that excluded instruments are exogenous, while underidentification test suggests that the model is identified. Although the instruments appear to be valid and exogenous, they are not very strong. I address the latter issue by using limited information maximum likelihood (LIML) estimator.

The IV estimation approach yields substantially lower estimates of the elasticity of substitution between leisure terms. This result would suggest that unobserved factors affect both wage and leisure gaps in the same direction. The new point estimates are now less than a half of that obtained by fixed effects, and the observed change in the values happens in the theoretically expected direction. Still, in all cases the point estimates of the elasticity of substitution are in excess of 1.5. The null of complementary leisure is rejected at 5% significance level or less in all three specifications, but only has a p-value below 0.01 in the final model that employs all instruments. These outcomes of leisure complementarity test are not as striking as in the baseline, yet the loss of precision may partly be due to the expected higher variance of the IV estimator. Nevertheless, even with the widest of all computed 95% confidence intervals, $-0.05 < \rho_L < 0.78$, we see that leisure complementarity, shall it be present, is extremely weak.

4.5. Home production

Another concern arises from the fact that the amount of leisure is computed as a simple difference between the fixed time endowment and the number of work hours, ignoring other aspects of time allocation and in particular home production. Household retirement decision is commonly considered in the context of time allocation between multiple uses that include work, leisure and home production (e.g., Rogerson and Wallenius, 2019; Stancanelli and Van Soest, 2012). As time allocated to home production changes endogenously in the course of transition from work to retirement, the increase of leisure associated with gradual retirement would be overstated. This may affect the household leisure gap and the outcomes of the leisure complementarity test.

It is straightforward to incorporate home production into the original household problem. Suppose that each spouse s can now allocate time endowment $\bar{L}$ between market work H_t^s, home production P_t^s and leisure L_t^s. The household can either purchase consumption good in the amount C_t^m in the market or produce it at home using production function $C_t^h = f(P_t^h, P_t^w)$ with decreasing marginal productivity. Assume that the utility from the overall consumption is given by the CES consumption subaggregate $C_t = \left[\alpha_C (C_t^m)^{\rho_C} + (1-\alpha_C)(C_t^h)^{\rho_C}\right]^{1/\rho_C}$, which is symmetric to the utility of leisure given by (2). This modification of the original model does not affect the ratio of the first order conditions for leisure terms (6), except that the amount of consumed leisure used in the estimation should now account for the household production, $L_t^s = \bar{L} - H_t^s - P_t^s$.

To estimate a version of the model with home production, I use data from the time use section of the HRS CAMS supplement. The supplement asks respondents about time spent on various activities over the reference period. I select eleven questions about activities relevant to home production. These include house cleaning and home improvements, gardening, shopping, preparing meals, unpaid work and helping others, and taking care of finances. I compute time each household spent in home production as a total value from these questions. This variable can then be used to estimate an equivalent of Eq. (7) with leisure gap accounting for time allocated to home production.

There are two challenges in the estimation of this model variation. First, the estimation sample is four times smaller than in the baseline because HRS CAMS is only administered to a sub-sample of chosen respondents. Second, it is impossible to estimate both nests of the utility function because the data do not identify separately consumption of home produced goods. Therefore, I can only estimate the inner nest of the household utility function on a smaller sample. The fixed effects estimates are consistent with the baseline case: the point estimate $\hat{\sigma}_L = 16.8$ is still very high and the null of leisure complementarity is rejected at all conventional significance levels. This estimate is likely to decrease with the use of instruments as it happened in the baseline case, yet based on the previous analysis there is no reason to expect that the decline would be so dramatic as to push the point estimate down below one. It does not seem plausible therefore that the definition of leisure variable is driving the test outcomes.

4.6. Discussion

To summarize the results, all model specifications yield point estimates of the elasticity of substitution above one, which is consistent with the hypothesis of substitutability between the leisure terms. In all cases, the null hypothesis of leisure complementarity is rejected at a significance level of 5% or lower. The preferred set of estimates from the model that uses all instruments (Model 4) rejects the hypothesis of leisure complementarity with a p-value of 0.009. Although the outcomes of the test of leisure complementarity are consistent throughout different model specifications, the point estimates of the elasticity of substitution parameter vary and may be affected by bias and measurement error. Two potential sources of bias, in particular, deserve to be mentioned.

First, an estimation bias can arise in the presence of corner solutions and contractual rigidities that restrict agents' ability to work as many hours as they would have liked without time constraints. Because the hours of leisure in Eq. (7) are a censored explanatory variable, the OLS and IV estimates of parameter β_{11} are likely affected by the expansion bias (Rigobon and Stoker, 2009). The expansion bias inflates estimates relative to the parameter values, and thus implies a true degree of substitutability higher than that suggested by the estimates. Therefore, this bias would further reinforce the outcomes of the complementarity test rather than refute it.

Second, a bias can arise from the model assumption that labor supply decisions do not affect future earnings and retirement income, in particular income from private pensions. To explore the magnitude and direction of this bias, I estimate the model using a sample of households that do not participate in any private pension plans and rely solely on Social Security for retirement income.[8] Unfortunately, such households constitute only 11% of the main estimation sample. Therefore, for this robustness check I relax sample restrictions related to missing data on the instruments and consumption. The fixed effects 95% confidence interval for the elasticity of substitution parameter in this sample, $0.84 < \rho_L < 0.99$, suggests that, similar to the previous case, the bias favors the reported outcome of the test of leisure complementarity.

Due to the potential bias, it is difficult to draw firm conclusions regarding specific point estimates of the elasticity of substitution. However, in both cases, the estimates of parameter ρ_L appear to be biased downwards, which means that eliminating the bias would further strengthen the results of the leisure complementarity test. Therefore, the overall compelling evidence from all estimated models suggests that, at best, the level of complementarity is negligibly low. More plausibly, the leisure of husband and wife in the household are gross substitutes.

Although these conclusions essentially disqualify leisure complementarity as a cause of retirement coordination, they are otherwise consistent with contemporary developments in the literature on household retirement. Rogerson and Wallenius (2019) show that older households appear to substitute leisure across their members. They estimate the

[8] This sample is based on the same approach as the sample of males in Rust and Phelan (1997), which aims to overcome complications arising from the lack of precise information about private pension plans.

Table 5
Calibrated parameters.

Parameter	Value			Source
	Household	Males	Females	
1. Preferences				
Leisure substitution within the household, ρ_L	0.39	–	–	IV estimates of the inner nest of the household utility function, Model 4 (Table 4)
Weight on husband's leisure, α_L	0.55	–	–	
Consumption/leisure substitution, ρ_C	0.95	0.99	0.99	Fixed effects estimates of the outer nest of the household/individual utility functions
Average weight on the household leisure, α_C	0.94	0.90	0.88	
Annual time trend in the leisure weight	–	0.01	0.03	Calibrated to match individual retirement rates
2. Mortality risk				
Age×0.01	–	−9.64	−10.1	Logit coefficients for biennial transition probabilities
Constant	–	9.01	9.69	
3. Wage transitions				
Constant	–	−4.28	−4.54	Conditional maximum likelihood estimates of Eq. (8)
Persistence of AR(1) component	–	0.97	0.98	
Variance of AR(1) component	–	0.07	0.07	
4. Other parameters				
Discounting factor, β	0.96			Values within the range used in the retirement literature (e.g. Scholz et al., 2006)
Return on assets, r	0.04			
Consumption minimum	9,228	–	–	Annual SSI payment for eligible couple in 2000

elasticity of substitution in leisure to lie in the interval between one and three. In spite of the methodological differences, this range of values is virtually the same as the 95% confidence interval of the estimates computed for Model 4 in this paper, $\sigma_L \in (1.06, 3.33)$. The results in Rogerson and Wallenius (2019) are based on a different identification strategy that exploits changes in time allocation between home production and leisure around the period of transition from full-time work to retirement. The advantage of my testing approach is in fewer requirements to the data. The test of leisure complementarity developed in this paper only requires data on labor supply and wages. Its results can easily be replicated using the basic variables commonly available in the standard labor force surveys even when detailed time use data are not available. A model extension with home production (Section 4.5) suggests that this simplification does not affect the estimates.

The idea that leisure complementarity motivates couples to retire together was further challenged by the lack of evidence that retired individuals spend more time in company of their partners. For example, Stancanelli and Van Soest (2016) show that in France time spent by couples in shared activities following retirement from the labor force is small relative to the increase in time allocated to individual activities. To add to this point, Eismann et al. (2017) use Dutch survey data to document that more dual career couples attribute their desire to coordinate retirement to the correlated preferences for leisure rather than preferences for sharing time together.

As I find that the co-movement of household wages and leisure choices in the data does not support the hypothesis of leisure complementarity, observed retirement coordination must emerge via a different channel. In the rest of the paper I explain how the model that was used to develop the test can generate retirement coordination in the absence of leisure complementarity.

5. The sources of retirement coordination

Simulation of the household retirement decisions in this section is based on the dynamic model of the household labor supply from Section 3. I solve the model by backward induction and simulate forward the path followed by each household. Simulated dataset contains a sequence of labor supply, consumption and saving decisions for 10,000 households recorded at annual time intervals. In accordance with the estimates obtained in the previous section, parameters of the household utility function do not allow for leisure complementarity. Nevertheless, simulated retirement pathways of older households exhibit a high degree of retirement coordination. I further show that a direct regression-based approach to the estimation of relationship between retirement decisions of the spouses, such as regression models discussed in Section 2, produces results similar to those reported in Table 2. Incorrect interpretation of the signs of regression coefficients in these models may lead to misleading conclusions about leisure complementarity.

5.1. Parametrization

Model parameters used in the simulations are summarized in Table 5. Section 1 of the table contains parameters of the household utility function. The inner CES nest, Eq. (2), is parameterized using preferred IV estimates (Model 4 in Table 4). The leisure substitutability parameter is set directly to the point estimate $\rho_L = 0.39$. The relative contribution of individual leisure terms to the utility function takes into account the estimated intercept and fixed effects and therefore varies by household; the average simulated weight $\alpha_L = 0.55$.

The outer nest is parameterized based on the fixed effects estimator discussed in Appendix C. Consumption and household leisure aggregate are estimated to be gross substitutes with $\rho = 0.95$. The weight α_t varies across households and time and involves estimated intercept, household fixed effects and a linear function in the ages of two household members. The age trend is calibrated so as to generate retirement rates consistent with the data and yields the overall weight on leisure of 0.94. Estimated parameters of single-member household utility functions are shown in gender-specific columns of the table.

The terminal individual age is set to $T = 100$. Individual survival probabilities are estimated by binary logit models conditional on age; their estimated coefficients are reported in Section 2 of Table 5. Individual wage transitions are modeled as error component processes with AR(1) disturbances; the estimates are reported in Section 3 of Table 5. The minimum consumption level is set at $C_{min} = \$9,228$. Discounting factor and the rate of return on assets are set at the values $\beta = 0.96$ and $r = 0.04$ as in Scholz et al. (2006).

The initial joint distribution of the state variables is drawn from the earliest available household observation in the HRS dataset. The data identify 1,309 types of households; the model is solved separately for each type. The number of simulated households of each type is determined by the HRS household sampling weights. Given the model solution and a sequence of randomly drawn shocks for the stochastic components of the state space, each simulated household selects consumption and labor supply decisions that maximize the expected lifetime utility, thus generating a simulated path. These simulations are used to generate counterfactual retirement scenarios and understand the process of retirement coordination.

Fig. 2. Model fit.
Notes: The top graphs show percentage of retired in the HRS data and simulations by age and gender. The bottom graphs show transition rates from work to retirement. Retirement status is defined by labor force participation. The HRS estimation sample includes coupled households from the HRS 2000–2016 with at least five years of job market experience; observations with missing data and incomplete retirement histories are excluded. The number of simulated households is $n_s = 10{,}000$. The shaded areas show 99% confidence bands.

5.2. Model fit

Retirement in the simulations is jointly generated by the Social Security incentives, age profiles in the individual survival probabilities and weights placed on aggregate leisure in the household decision making. The latter is calibrated to match individual retirement rates by age and gender in the data. I use Nelder–Mead simplex search algorithm to find the weight on household leisure and age trend in the leisure preferences that minimize the distance between simulated and observed individual retirement rates. The average leisure weight for households with age of both spouses set to 55 is 0.93; over the next ten years of life it increases to 0.95. The model yields the average weight on household leisure of 0.94 that is consistent with reduced form estimates of the utility function.

Gender differences in the individual age trends make women retire at younger ages than men. Although this helps match the retirement data, it is beyond the scope of this paper to answer what creates such differences. One possible mechanism that is not taken into account by the model is health shocks and gender differences in the caretaking responsibilities that can make females retire at younger age. Another one is subjective reference points that play an important role in retirement decisions (Seibold, 2021). Any perceived gender differences in the reference points will be absorbed by the leisure weights in this model. Gender differences in involuntary unemployment and job search outcomes at older age will also be captured by the age trends.

To evaluate the quality of model fit, Fig. 2 compares simulated age-specific male and female retirement rates, which were used as targeted moments in the calibration, to the data. Retirement rates computed from the simulations show gradual transition from work to retirement that is consistent with the data. Most of the simulated values are contained within the gray area that shows the 99% confidence band of the data estimates.

Unlike individual retirement rates, the distribution of differences in the retirement dates of the household members and the degree of retirement coordination are not targeted in the calibration. Any pattern in joint retirement therefore emerges as an outcome of the model. I validate the model by comparison of its predictions to the descriptive data facts that were introduced in Section 2. In 34.5% of all simulated couples, the spouses retire within one year of each other. This result corresponds to the 34% of couples in the estimation sample that reported retirement in the same year.

Fig. 3, a simulated analogue of Fig. 1, plots the distribution of within household differences in retirement dates generated by the model. The solid line is the kernel density of within couple differences in the simulated retirement dates. For comparison, the distribution of age differences in the couples that is drawn from the data as a part of initial state is shown by the dashed line. Much the same as the distribution observed in the data, the distribution of differences in retirement dates is more symmetric around zero than that of the age differences. Overall, the model generates substantial levels of retirement coordination that are consistent with the patterns observed in the data.

Suppose now that we tried to make inference about retirement coordination and leisure complementarity by regressing individual retirement status on the retirement status of the spouse. In the simulated data, the probability of being out of labor force is respectively 0.168 and 0.173 higher for males and females when their spouse is retired. The data counterparts of these numbers, which are the coefficients in the interdependent retirement equations 0.167 and 0.171 in Table 2, provide additional validation of the model. Interpreting the positive coefficient sign in the estimated relationship as a link to leisure complementarity in this case would be incorrect since we know that there is no complementarity in the underlying household utility function. In the meantime, fixed effects estimates of Eq. (6) correctly infer the elasticity of substitution between individual leisure terms to be above one.

5.3. The sources of retirement coordination

The model allows to identify three channels that can entice households to coordinate their retirement decisions: shared budget constraint, policy incentives and the distribution of preferences for leisure. In the rest of the paper, I estimate the extent to which each of these channels can account for observed retirement coordination, and quantify their relative importance. I do it by running a set of counterfactuals in which households are identical to those in the original simulated data in terms of all initial state characteristics and stochastic shocks received throughout the lifetime, except for selected model features or policy incentives that potentially account for some of retirement coordination. The results reveal how the engagement of each of the potential

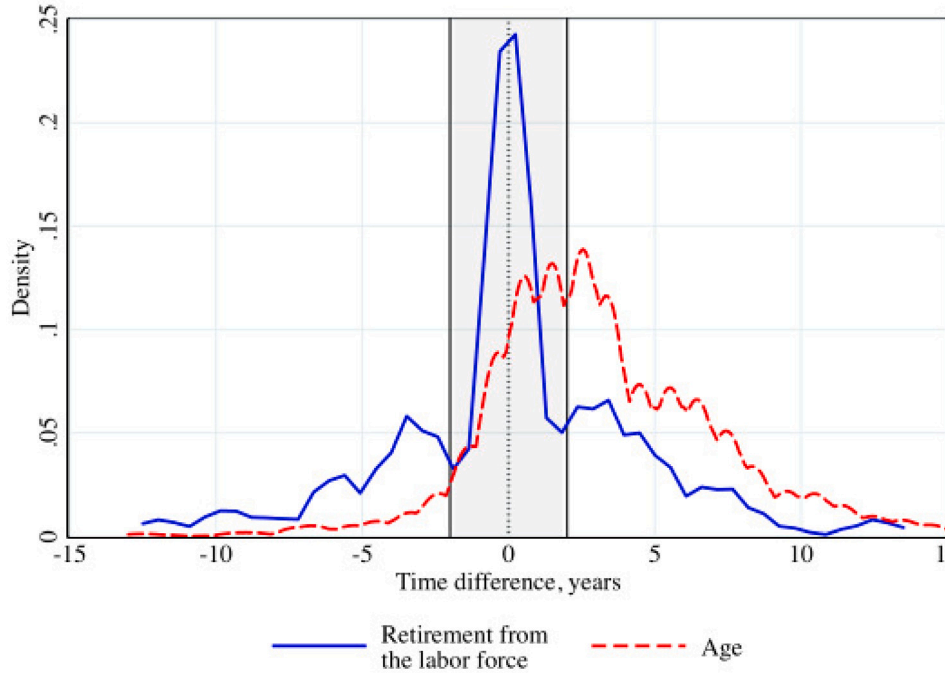

Fig. 3. Retirement coordination: simulations.
Notes: The plots show density functions estimated with Epanechnikov kernel and bandwidth $h = 0.5$. The time differences are measured as male minus female outcome in years; all differences are truncated at 15 years. Retirement status is defined by labor force participation; retirement date is the end date of the last job held prior to the first-time exit from the labor force. The vertical dotted line corresponds to zero time difference; the shaded area is the time frame of two years either side of zero.

coordination channels affects the degree of retirement coordination in the simulations.

To facilitate the discussion, I rely on the fraction of couples retired within one year of each other as the main measure of retirement coordination. The benchmark values of this indicator in the simulated data is 34.5% of the couples that retired together. The average individual age of retirement from the labor force measures the effect of proposed sources of retirement coordination mechanisms on the individual retirement decisions; the baseline simulated values are 63.2 years for males and 59.9 for females. Table 6 summarizes how these measures change in the counterfactuals. It also shows how households arrive at new coordination outcomes by computing the adjustment margins—the percent of counterfactual households with changed coordination outcome where the larger adjustment was made by either male or female, or both household members equally.

The relationship between household retirement coordination and the budget constraint is illustrated by two counterfactuals. First, the household retirement plans may be affected by events that alter the budget constraint shared by both of its members. For a specific example, consider a scenario where the total household wage income is decreased by 20% from the time the oldest household member reaches the age of 60 onwards. This shock equally affects both household members and may have a variety of interpretations such as job market shocks or unanticipated large expenses due to health and caretaking. As a result of this shock, male and female household members postpone their retirement by 2.4 and 3.6 months, respectively. While on average retirement is postponed, the overall degree of coordination gets higher because women delay their retirement more: the fraction of couples to retire within one year of each other increases by 16.5%. Retirement coordination therefore is linked to shared shocks to household income and wealth: as both household members fall under pressure to work longer in order to offset the negative income of a shock, they eventually retire closer to each other.

Notice that the model does not take into account health and medical expenditures. This results in the underestimation of the amount of assets held at the end of life: in the absence of catastrophic medical expenses, as well as long-term care costs and bequest motives, households use up their savings faster than they do in the data. The results of income and wealth counterfactuals should, therefore, be regarded as conservative estimates. However, the retirement incentives related to the state of health and healthcare costs do not seem as important for the results on employment rates and coordination at the stage of transition from work to retirement. In particular, households where both members are older than 65 and eligible for Medicare and younger households that are more likely to rely on employer-provided health insurance both exhibit behavior that is consistent with leisure substitutability (there is no statistically significant difference in the estimated complementarity parameter ρ_L between these two groups, p-value for the H_0 of equal coefficients of 0.642). Furthermore, although medical expenditures, health insurance, and Medicare are important for understanding retirement behavior, quantitatively their effects on employment are not very large. For example, French and Jones (2011) compute that two additional years without access to Medicare increase the duration of working lives by less than one month.

The second type of shock affecting the budget constraint, such as for example individual job loss, would have a differential effect on the household members. I quantify the effect of such shocks by setting the wage rate to zero around the average retirement time of the affected individual: 63 for males and 60 for females. The shock results in immediate forced retirement of the affected household member. Male job loss in particular visibly moves forward the retirement of the partner, although marginally so, same as in the case with total income shock: by less than a quarter of a year. The number of coordinated retirements falls by approximately 30%, suggesting that forced retirement of the partner does not provide additional incentives for immediate own withdrawal from the labor force. On the contrary, a response akin to the added worker effect keeps the employed partner working.

The next channel that can generate retirement coordination is Social Security policy which also largely operates through the household

Table 6
Retirement coordination in the counterfactuals.

Model	Average retirement age, years		Percent retired in the same year	Adjustment margin, % of households with changed coordination outcome		
	Males	Females		Male	Female	Both
Baseline simulation	63.2	59.9	34.5			
Decrease in the household income, 20%	63.4	60.2	40.2	36.0	49.9	14.1
Loss of individual wage income:						
Husband	61.9	60.1	23.5	92.2	7.8	0.0
Wife	63.2	58.8	23.8	4.0	92.5	3.5
Social security policy:						
Increase eligibility age by 1 year	63.4	60.3	35.5	44.8	53.5	1.7
Reduce benefits by 20%	63.4	60.2	37.5	26.9	54.4	18.7
Eliminate family benefits	63.5	60.4	33.8	28.7	63.0	18.3
Increase weight on household leisure by 10%	62.6	57.9	30.7	0	0	100
Increase weight on husband's leisure by 10%	63.0	60.4	38.8	33.6	50.8	15.6
Eliminate household wage gap	63.6	61.4	45.6	11.0	76.4	12.6
Reduce age difference by one year	63.1	60.5	34.7	31.2	43.9	24.9

Notes: Household income shock counterfactual is for 20% loss in wages by both household members when the oldest turns 60. Individual loss of wage income happens at the average age of retirement in the baseline simulations: 63 for males and 60 for females. Household wage gap is reduced by assuming the same (male) parameters of the wage determination process for both genders. The adjustment margins give percent of counterfactual households with changed coordination outcome where the larger adjustment is made by either male or female, or both household members equally.

budget constraint, albeit the source of possible shocks is quite different. I consider three counterfactuals: increase in eligibility age, reduction of benefits, and elimination of spousal benefits. All three experiments affect the household retirement in the expected direction that is consistent with previously discussed results. The effects however are much smaller in magnitude than those obtained earlier for the shocks to the budget constraint. It is possible that the relationship between Social Security policy and retirement is dampened by the model assumptions limiting the impact of labor supply decisions on the retirement wealth accruals. The results that follow should therefore be regarded as conservative estimates of policy changes on individual retirement decisions. However because both household members are affected by these assumptions in similar way, it is unlikely that relaxing the assumptions will change the main conclusions about retirement coordination.

A one-year increase to both early and full retirement age delays individual retirement by on average 2.5 months for males and one third of a year for females. It slightly increases retirement coordination, by approximately one percentage point. The impact of reduction in the overall amount of benefits is less straightforward as it affects both the household income and the incentives to increase benefit amounts by working longer. Overall, 20% reduction in the individual benefits increases the fraction of coordinated retirements to 37.5%. Qualitatively, both policy changes have an effect similar to that of decrease in the overall household income.

The Social Security rules allow individuals to choose between benefits that are based on their own earnings and up to a half of their spouse's benefits, whichever is higher. This option links retirement incentives within a couple: household members with low lifetime earnings may plan to retire together with their partners and claim their benefits. In the following counterfactual, I assume that this option is not available, so that each individual is only eligible for own retirement benefits. I also eliminate the rules that restrict the maximum amount of benefits available to a household. As a result, the number of coordinated retirements in the simulations decreases by 0.7 percentage points. This change in the retirement coordination takes place as females increase their labor force participation while males retire earlier, a pattern qualitatively similar to that established by Borella et al. (2023). The effects are likely smaller in magnitude because the model in this paper does not take into account the role of taxes; there are also more limited saving motives so that the households are able to adjust consumptions by divesting of their assets instead of adjusting their labor supply. Overall, the estimated impact of family provisions on retirement coordination is positive, but small in comparison to that of other channels.

The last channel of retirement coordination is represented by the household preferences, including the overall household taste for leisure and the relative weights on the individual leisure terms in the utility function. A 10% increase in the weight on the household leisure composite (parameter α_t in Eq. (3)) decreases retirement coordination by 11%. Therefore, retirement coordination is strongly related to the factors that shift the weight on the household leisure composite in the utility function. An increase in the overall importance of leisure over consumption in the household decision making would make each partner reduce own labor supply, and lead to more synchronized retirement. This scenario can cover the role of assortative matching in marriage. Couples match on many factors, possibly including similar tastes for leisure. If this is the case, we would observe coordinated retirement simply because of implicitly shared understanding of the right time to leave the labor force. Survey results in Eismann et al. (2017) support this reasoning.

The relative weights placed on the leisure of individual household members are also important for the coordination outcome. A 10% increase on the weight placed on husband's leisure (parameter α_L in Eq. (2)) generates 12.5% increase in retirement coordination. It appears that more coordination would be observed when individuals offer more even contribution to the household utility as measured by their leisure weights and potential earnings. In the absence of gender wage gap within households, coordination would increase by 32%, more than in any other counterfactual discussed so far.

Finally, I look into the role of age differences that was initially discussed in Section 2. I decrease the age difference by taking away one year from the age of the older partner in couples with age difference of one or more years. With lower average age difference, the fraction of couples to retire within the same period increases slightly to 34.7%. It does not seem that age distribution is essential for the process of retirement coordination.

A common pattern emerging from all counterfactuals is that instead of coordinating retirement because of the value placed on time spent together at home, couples with more equal positions of spouses in terms of earned wages, tastes for leisure and attachment to the labor force are more likely to retire together. In addition, females exhibit stronger response to the shocks, and so coordination outcomes are more likely to change due to the female retirement adjustment. An implication is that we would expect to see higher retirement coordination both with the increase of female labor force participation rates and the closing of the gender wage gap. This explains the persistence of retirement coordination in the data that is observed regardless, or rather as it appears due to, the structural changes in the employment patterns of older households.

6. Conclusions

This paper uses the data from the Health and Retirement Study to test whether complementarity of leisure in dual career households can explain coordinated retirement from the labor force. I develop a test of leisure complementarity that is based on a dynamic model of household labor supply with flexible CES preferences. My estimates show that leisure terms of the spouses in the household utility function are strong substitutes rather than complements.

Having shown that leisure complementarity does not seem to account for observed retirement coordination, I turn to other possible explanations. Using estimated parameters of the household utility function, I calibrate a dynamic programming model with uncertainty about household survival and wage earnings. The extent of retirement coordination generated by the model conforms with the data, and overall the model captures well the household transition from work to retirement. I further use a set of counterfactuals to evaluate the role that earnings shocks, Social Security policy and household tastes for leisure each play in retirement coordination.

Instead of leisure complementarity, retirement coordination in the model is linked to the degree of similarity between the profiles of the household members, or lack thereof. Given the structure of the household preferences and the process of asset accumulation, retirement starts once the weight placed on the aggregate leisure consumed by the household becomes sufficiently high relative to that placed on consumption. In a household with symmetric leisure preferences and two identical household members both would follow the same retirement path and exit the labor force around the same time, giving an impression of coordinated decision. Otherwise, a household would optimize its lifetime utility by retiring the member with higher contribution to the aggregate leisure first. Therefore, the degree of retirement coordination depends on within household differences in wages, lifetime earnings and weights on individual leisure. Equalizing the lifetime earnings of the two household members by fully closing the earnings gap would bring the number of couples retiring in coordinated fashion close to one half.

These result are important because leisure complementarity is often referenced in the retirement literature as a routine explanation of retirement coordination. Knowing the relationship between the leisure of husbands and wives can benefit policy makers, as the joint household response to policy measures will depend on the interaction of leisure and consumption terms in the household utility function. For example, if the leisure terms were complementary, we could expect a magnified response to gender specific policies. This will not be the case for substitutable leisure terms and coordination that arises from the similarities of individual contribution to the household utility from the overall leisure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.labeco.2023.102431.

References

An, M.Y., Christensen, B.J., Gupta, N.D., 2004. Multivariate mixed proportional hazard modelling of the joint retirement of married couples. J. Appl. Econometrics 19 (6), 687–704.
Atalay, K., Barrett, G.F., Siminski, P., 2019. Pension incentives and the joint retirement of couples: Evidence from two natural experiments. J. Popul. Econ. 32, 735–767.
Banks, J., Blundell, R., Casanova Rivas, M., 2010. The dynamics of retirement behavior in couples: reduced-form evidence from England and the US. Manuscript.
Bingley, P., Lanot, G., 2007. Public pension programmes and the retirement of married couples in Denmark. J. Public Econ. 91 (10), 1878–1901.
Blau, D.M., 1998. Labor force dynamics of older married couples. J. Labor Econ. 16 (3), 595–629.
Borella, M., De Nardi, M., Yang, F., 2023. Are marriage-related taxes and Social Security benefits holding back female labour supply? Rev. Econom. Stud. 90 (1), 102–131.
Browning, M., Deaton, A., Irish, M., 1985. A profitable approach to labor supply and commodity demands over the life-cycle. Econometrica 53 (3), 503–543.
Coile, C., 2004. Retirement incentives and couples' retirement decisions. Top. Econ. Anal. Policy 4, 1–28.
Eismann, M., Henkens, K., Kalmijn, M., 2017. Spousal preferences for joint retirement: Evidence from a multiactor survey among older dual-earner couples. Psychol. Aging 32 (8), 689–697.
French, E.B., Jones, J.B., 2011. The effects of health insurance and self-insurance on retirement behavior. Econometrica 79 (3), 693–732.
Gustafson, P., 2017. Spousal age differences and synchronised retirement. Ageing Soc. 37 (4), 777–803.
Gustman, A.L., Steinmeier, T.L., 2000. Retirement in dual-career families: A structural model. J. Labor Econ. 18 (3), 503–545.
Heckman, J.J., Lochner, L., Taber, C., 1998. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. Rev. Econ. Dyn. 1 (1), 1–58.
Hospido, L., Zamarro, G., 2014. Retirement patterns of couples in Europe. IZA J. Eur. Labor Stud. 3 (1), 1–18.
HRS, 1992-2014. Health and retirement study public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2017).
Hurd, M.D., 1990. The joint retirement decision of husbands and wives. In: Wise, D.A. (Ed.), Issues in the Economics of Aging. University of Chicago Press, pp. 231–258.
Kapur, K., Rogowski, J., 2007. The role of health insurance in joint retirement among married couples. ILR Rev. 60 (3), 397–407.
Kniesner, T.J., 1976. An indirect test of complementarity in a family labor supply model. Econometrica 44 (4), 651–669.
Knowles, J.A., 2013. Why are married men working so much? An aggregate analysis of intra-household bargaining and labour supply. Rev. Econom. Stud. 80 (3), 1055–1085.
Lalive, R., Parrotta, P., 2017. How does pension eligibility affect labor supply in couples? Labour Econ. 46, 177–188.
Mastrogiacomo, M., Alessie, R., Lindeboom, M., 2004. Retirement behaviour of Dutch elderly households. J. Appl. Econometrics 19 (6), 777–793.
Michaud, P.-C., Van Soest, A., Bissonnette, L., 2020. Understanding joint retirement. J. Econ. Behav. Organ. 173, 386–401.
Michaud, P.-C., Vermeulen, F., 2011. A collective labor supply model with complementarities in leisure: Identification and estimation by means of panel data. Labour Econ. 18 (2), 159–167.
Rigobon, R., Stoker, T.M., 2009. Bias from censored regressors. J. Bus. Econom. Statist. 27 (3), 340–353.
Rogerson, R., Wallenius, J., 2019. Household time use among older couples: Evidence and implications for labor supply parameters. Q. J. Econ. 134 (2), 1079–1120.
Rust, J., Phelan, C., 1997. How Social Security and Medicare affect retirement behavior in a world of incomplete markets. Econometrica 65 (4), 781–831.
Schirle, T., 2008. Why have the labor force participation rates of older men increased since the mid-1990s? J. Labor Econ. 26 (4), 549–594.
Scholz, J.K., Seshadri, A., Khitatrakun, S., 2006. Are Americans saving "optimally" for retirement? J. Polit. Econ. 114 (4), 607–643.
Seibold, A., 2021. Reference points for retirement behavior: Evidence from German pension discontinuities. Amer. Econ. Rev. 111 (4), 1126–1165.
Stancanelli, E., 2017. Couples' retirement under individual pension design: A regression discontinuity study for France. Labour Econ. 49 (2017), 14–26.
Stancanelli, E., Van Soest, A., 2012. Retirement and home production: A regression discontinuity approach. Am. Econ. Rev. 102 (3), 600–605.
Stancanelli, E., Van Soest, A., 2016. Partners' leisure time truly together upon retirement. IZA J. Labor Policy 5 (1), 5–12.
van der Klaauw, W., Wolpin, K.I., 2008. Social Security and the retirement and savings behavior of low-income households. J. Econometrics 145 (1-2), 21–42.
Warren, D.A., 2015. Retirement decisions of couples in Australia: The impact of spousal characteristics and preferences. J. Econ. Ageing 6, 149–162.