cemmap

centre for microdata methods and practice

A new model for interdependent durations with an application to joint retirement

Bo E. Honoré
Áureo de Paula

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP07/16

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

A New Model for Interdependent Durations with an Application to Joint Retirement*

Bo E. Honoré †

Princeton University

Áureo de Paula ‡

UCL, São Paulo School of Economics-FGV, C*e*MMAP, IFS

This Version: February 2016

*We thank Manuel Arellano, Richard Blundell, Stephane Bonhomme, Martin Browning, Ken Chay, Liran Einav, Eric French, Han Hong, Rafael Repullo, Geert Ridder, Jim Smith, Gerard van den Berg and Frank Wolak for useful conversations and David Card for suggesting this application. We are also grateful to the editor, Petra Todd, and three anonymous referees, who provided many remarks and comments that helped us improve the article. Financial support from a Steven H. Sandell grant from the Center for Retirement Research (Boston College) is gratefully acknowledged. Bo Honoré also acknowledges financial support from the National Science Foundation and from the Gregory C. Chow Econometric Research Program. Áureo de Paula also acknowledges financial support from the National Science Foundation, the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 and the European Research Council through the ERC Starting Grant 338187. Andrea Puig and Reka Zempleni provided very able research assistance.

†Princeton University (Department of Economics), Princeton, NJ 08542. E-mail: `honore@princeton.edu`

‡University College London (Department of Economics), São Paulo School of Economics-FGV, C*e*MMAP and IFS, London, UK. E-mail: `apaula@ucl.ac.uk`

Abstract

This paper introduces a bivariate version of the generalized accelerated failure time model. It allows for simultaneity in the econometric sense that the two realized outcomes depend structurally on each other. Another feature of the proposed model is that it will generate equal durations with positive probability. The motivating example is retirement decisions by married couples. In that example it seems reasonable to allow for the possibility that each partner's optimal retirement time depends on the retirement time of the spouse. Moreover, the data suggest that the wife and the husband retire at the same time for a non-negligible fraction of couples. Our approach takes as a starting point a stylized economic model that leads to a univariate generalized accelerated failure time model. The covariates of that generalized accelerated failure time model act as utility-flow shifters in the economic model. We introduce simultaneity by allowing the utility flow in retirement to depend on the retirement status of the spouse. The econometric model is then completed by assuming that the observed outcome is the Nash bargaining solution in that simple economic model. The advantage of this approach is that it includes independent realizations from the generalized accelerated failure time model as a special case, and deviations from this special case can be given an economic interpretation. We illustrate the model by studying the joint retirement decisions in married couples using the Health and Retirement Study. We provide a discussion of relevant identifying variation and estimate our model using indirect inference. The main empirical finding is that the simultaneity seems economically important. In our preferred specification the indirect utility associated with being retired increases by approximately 5% when one's spouse retires. The estimated model also predicts that the marginal effect of a change in the husbands' pension plan on wives' retirement dates is about 3.3% of the direct effect on the husbands'.
JEL Codes: J26, C41, C3.

1 Introduction and Related Literature

This paper introduces a new class of econometric duration models that allow for joint determination of pairs of durations. This joint determination can manifest itself not only in correlation between the durations, but also in a non-zero probability that they are equal, despite their marginal distributions being continuous. The class has the generalized accelerated failure time model as a special case.

One easy way to introduce correlation in two durations is to allow for correlation in unobserved heterogeneity. In addition, it would be easy to allow for concurrent exits by allowing for common shocks in the spirit of Marshall and Olkin (1967). In contrast to this, the aim of this paper is to introduce a model in which the dependence is generated endogenously. This is in many ways similar to introducing the correlation in a pair of linear regressions through simultaneity as opposed to through correlation in the errors.

Our approach is to think about one individual in a pair making a transition into a new state, where the utility in the new state depends on whether the other individual in the pair is in the state. These utility externalities are in the spirit of de Paula (2009) and Honoré and de Paula (2010). Those papers assume an environment where a non-cooperative model is natural. However, when the utility-externality is positive and the individuals can communicate, it may be more reasonable to model the observed transition times as the outcomes of a Nash bargaining problem. In Section 2 of this paper, we set up a stylized model that has this flavor. In doing so, we will make a number of admittedly restrictive assumptions. Those assumptions are all driven by the goal of keeping standard continuous-time econometric duration models as special cases of our model.

The paper applies the econometric model to joint retirement decisions within married couples. A majority of retirees are married and many studies indicate that a significant proportion of individuals retire within a year of their spouse. Articles documenting the joint retirement of couples (and the datasets employed) include Hurd (1990) (New Beneficiary Survey); Blau (1998) (Retirement History Study); Gustman and Steinmeier (2000) (National Longitudinal Survey of Mature Women); Michaud (2003) and Gustman and Steinmeier (2004) (Health and Retirement Study); and Banks, Blundell, and Casanova Rivas (2007) (English Longitudinal Study of Ageing). Even though this is especially the case for couples closer in age, a spike in the distribution of retirement time differences at zero typically exists for most couples, regardless of the age difference. This is illustrated in Figure 1.

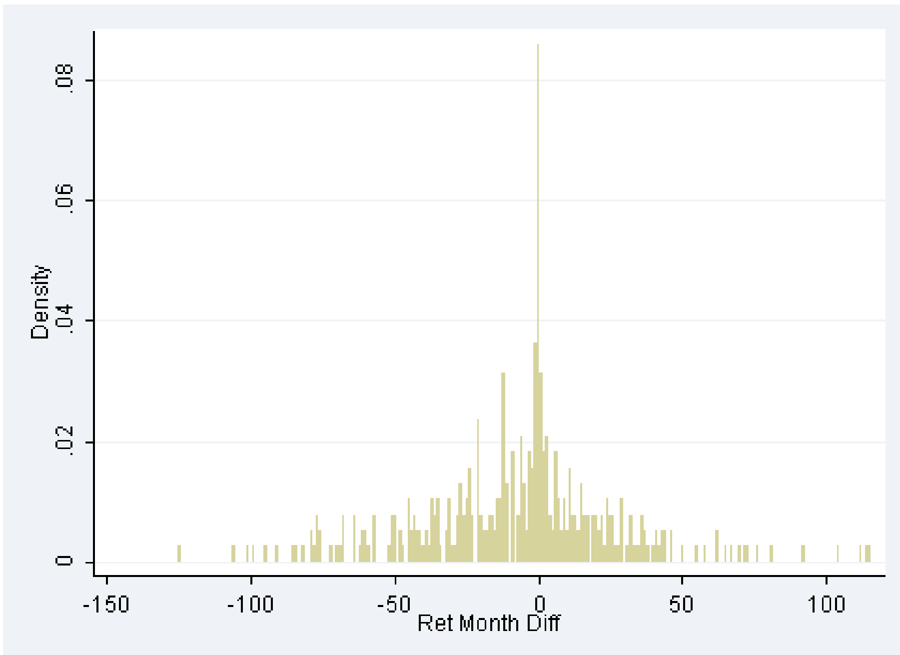

Figure 1: Difference in Retirement Months (Husband-Wife)

The spike in the distribution of the difference in retirement dates for husbands and wives in Figure 1 suggests that many couples retire simultaneously. This is consistent with the observation that 55% of respondents in the Health and Retirement Study expected to retire at the same time as their spouses.[1] There are at least two explanations for such a phenomenon. One is that the husband and wife expect to receive correlated shocks (observable or not), driving them to retirement at similar times. This is similar to a Marshall and Olkin (1967) model, and it is the approach used by An, Christensen, and Gupta (2004) to analyze joint retirement in Denmark. The other explanation is that retirement is jointly

[1]The figure corresponds to those who answer either YES or NO to the question: "Do you expect your spouse to retire at about the same time that you do?" (R1RETSWP). It excludes those whose spouse was not working.

decided, reflecting the taste interactions of both members of the couple. In this paper, we focus on the second of these explanations because it is this mechanism that corresponds to our methodological contribution.

The distinction between the two drivers of joint retirement (which are not mutually exclusive) is similar to the motivation for studying linear simultaneous equation models, and it parallels the categorization by Manski (1993) (see also de Paula (2016)) of correlated and endogenous (direct) effects in social interactions. In those literatures, the joint determination of the outcomes of interest y_i for individuals $i = 1, 2$ is represented by the system of equations

$$\begin{aligned} y_1 &= \alpha_1 y_2 + x_1' \beta_1 + \varepsilon_1 \\ y_2 &= \alpha_2 y_1 + x_2' \beta_2 + \varepsilon_2, \end{aligned}$$

where x_i and $\varepsilon_i, i = 1, 2$ represent observed and unobserved covariates determining y_i. We want to separate the endogenous (direct) effect (the α's) from the correlation in the ε's. Discerning these two sources of correlation in outcomes is relevant for analytical and policy reasons. For example, when the estimated model does not allow for joint decision making within the couple, the estimate of the effect of a retirement-inducing policy shock can be misleading if the retirement dates are indeed chosen jointly. The spillover effects that result from joint decisions invalidate, for instance, the commonly employed Stable Unit Treatment Value Assumption used in the treatment effects literature. This prevents a clear separation of the direct effects and the indirect effects that occur through feedback to the partner's retirement decision (e.g., Burtless (1990)). Furthermore, the multiplier effect induced by the effect of one person's retirement on the spouse is a potentially important conduit for policy. The quantification of its relative importance is therefore paramount for both methodological and substantive reasons.

The broader literature on retirement is abundant, and a number of papers focusing on retirement decisions in a multi-person household have appeared over the last 20 years. Hurd (1990) presents one of the early documentations of the joint retirement phenomenon. Later papers confirming the phenomenon and further characterizing the correlates of joint retirement are Blau (1998); Michaud (2003); Coile (2004a); and Banks, Blundell, and Casanova Rivas (2007). Gustman and Steinmeier (2000) and Gustman and Steinmeier (2004) work with a dynamic economic model in which the husband's and wife's preferences are affected by their spouse's actions, but the couple makes retirement decisions individually.[2] These papers focus on Nash equilibria to the joint retirement decision, i.e., each spouse's retirement decision is optimal given the other spouse's timing and vice-versa.[3] More recently, Gustman and Steinmeier (2009) present a richer (non-unitary) economic model with a solution concept that differs from a Nash equilibrium and is guaranteed to exist and be unique. Michaud and Vermeulen (2011) estimate a version of the "collective" model introduced by Chiappori (1992) in which (static) labor force participation decisions by husband and wife are repeatedly observed from a panel (i.e., the Health and Retirement Study). Casanova Rivas (2010) suggests a detailed unitary dynamic economic model of joint retirement. Coile (2004b) presents statistical evidence on health shocks and couples' retirement decisions and Blau and Gilleskie (2004) present an economic model that also focuses on health outcomes and couples' retirement decisions.

In applying our econometric approach to joint retirement, we implicitly assume that retirement decisions are made through Nash bargaining on the retirement date. This solution concept is attributed to Nash (1950) (see also Zeuthen (1930)). It chooses retirement decisions to maximize the product of differences between spouses' utilities and respective outside options (i.e., "threat-points"). The Nash solution corresponds to a set of behavioral axioms on the bargaining outcomes (essentially Pareto efficiency, independence of irrelevant alternatives and symmetry), and it is widely adopted in the literature on intra-household

---

[2]In the family economics terminology, their model is a non-unitary model in which people in the household make decisions individually. In unitary models, the household is viewed as a single decision-making unit. A characterization of unitary and non-unitary models can be found in Browning, Chiappori, and Lechene (2006).

[3]When more than one solution is possible, they select the Pareto dominant equilibrium, i.e., for all other equilibria at least one spouse would be worse off. If no equilibrium is Pareto dominant, the equilibrium where retirement by at least one household member happens earliest is assumed (see, e.g., Gustman and Steinmeier (2000), pp. 515, 520).

bargaining. It can be shown that this solution approximates the equilibrium outcome of a situation in which husband and wife make offers to each other in an alternating order, and the negotiation breaks down with a certain probability. As this probability goes to zero, the equilibrium converges to the Nash solution (see Binmore, Rubinstein, and Wolinsky (1986)). Though this solution also leads to Pareto efficient outcomes, it imposes more structure than Casanova Rivas (2010) or Michaud and Vermeulen (2011) (see Chiappori (1992) and Chiappori, Donni, and Komunjer (2012)).

Our econometric model is a variation of a recently developed model (Honoré and de Paula (2010)) that extends well-known duration models to a (non-cooperative) strategic stopping game, in which endogenous and correlated effects can be disentangled and interpreted (see also de Paula (2009) for a related analysis). However, our model extends simultaneous duration models differently from Honoré and de Paula (2010): whereas that paper suggests a non-cooperative game theoretic framework, the use of a cooperative framework is much more appealing for applications where the utility-externality is positive and where individuals can cooperate. Like the framework in Honoré and de Paula (2010), the model proposed here directly corresponds to an economic model of decision-making, and it can consequently be more easily interpreted in light of such a model. To estimate our model, we resort to indirect inference (Smith (1993); Gourieroux, Monfort, and Renault (1993); and Gallant and Tauchen (1996)), using as auxiliary models standard duration models and ordered discrete choice models, as suggested in Honoré and de Paula (2010) for a similar framework. (For an earlier application of indirect inference in a duration context, see Magnac, Robin, and Visser (1995)).

The remainder of this paper proceeds as follows. Section 2 describes our model and the empirical strategy for its estimation. In Section 3 we briefly describe the data and subsequently discuss our results in Section 4. We conclude in Section 5.

2 Model and Empirical Strategy

2.1 Basic Setup

In this section we formulate a simple econometric framework that allows a pair of durations with continuous marginal distributions to be interdependent and equal with positive probability. We discuss the model in the context of retirement decisions within a household, but it can be applied to any context in which the exit times from an initial state to the destination state are chosen optimally, and in which it may be optimal that the exit times are coordinated. Our strategy is to model this in the spirit of a discrete choice model in which the individual compares the utility in the two states. The interdependence is driven by the possibility that the utility flow in the destination state (retirement) depends on whether the other person is already in the state. In the case of retirement decisions, this captures the idea that spouses will want to decide jointly when to retire, and that the optimal decision can be to retire at the same time if the utility flow from retirement depends on the retirement status of the spouse. As is usual in choice models, the choice of the transition times depends only on the difference in the discounted future utilities between being in the initial state and being in the new state. The levels of the utilities do not matter. This implies that many of the seemingly arbitrary assumptions made below are mere normalizations with no behavioral implications.

The resulting econometric model is explicitly designed to have the proportional hazard model as a special case. In this sense, it is a true generalization of a standard econometric model.

To fix ideas, it is useful to introduce the econometric framework in the context of the empirical example. However, we emphasize that the setup can be useful in any situation in which two or more durations are coordinated. In our model, a pair of individuals i and j each choose when to transition from an initial state (in our case working) to a destination state (in our case retirement). i and j will take values 1 and 2.

Individual i with observable characteristics, $\mathbf{x}_i$, receives a utility flow of $K_i > 0$ in the initial state (in the example, working). In the destination state, the utility flow at time s is given by the deterministic function, $H_i(s, \mathbf{x}_i) D(s, t_j)$ where t_j is the time at which j transitions. The function $D(s, t_j)$ is defined as $(\delta - 1)1(s \geq t_j) + 1$ with $\delta \geq 1$ and it captures the idea that there can be complementarities in the transition decisions; the utility for i in the destination state is higher once j has made the transition. In the empirical example, the utility of being retired is higher if the spouse is also retired. The complementarities implied by $D(s, t_j)$ can be ascribed either to taste or to institutional features. In the retirement example these include tax or Social Security rules that may promote coordination in retirement timing between husband and wife. Whereas this parameter would not be invariant to changes in such regulations, it may be taken as fixed with respect to other counterfactuals. The parameter δ could in principle be less than one. However, this would not generate a positive probability that the individuals retire at the same time (as observed in the data). In the calculations and exposition below, we therefore restrict our attention to the case where δ is greater than or equal to 1. The δ could be made spouse-specific as well, but we focus on homogeneous δ. The reason for this is simplicity, and the fact that while the probability of joint retirement will be driven by δ, it is difficult to think of features of the data that would allow us to separately identify a different δ for husbands and for wives.

The function $H_i(s, \mathbf{x}_i)$ is assumed to be increasing in s. This is because we are interested in a single spell econometric model in which each individual makes one transition. In the application, this makes retirement an absorbing state. In the general discussion below, the key feature of $H_i(s, \mathbf{x}_i)$ is that its path is known at the time when the transition decision is made and that it is increasing.[4] In the empirical application, we will assume that it is separable, but this is not necessary. Moreover, the covariates can be time-varying, in which case $\mathbf{x}_i$ denotes the time-path of the explanatory variable, and $H_i(s, \mathbf{x}_i(s))$ is then assumed to be increasing.

As mentioned above, only the difference in utilities matters. This means that in the

[4]Without the monotonicity, retirement would not be an absorbing state, which would complicate the analysis.

retirement example, the monotonicity assumption implies that retirement becomes relatively more attractive over time. Of course, this does not imply that some absolute measure of happiness increases with age. The multiplicative structure for $H_i(s, \mathbf{x}_i) D(s, t_j)$ is imposed because we want the resulting model to have the same structure as the familiar proportional hazard model. Except for this, the functional form for the utility flow could easily be relaxed. In principle, it is possible to allow for kinks or discontinuities in $H_i(\cdot, \mathbf{x}_i)$. In a model without interdependence, those would correspond to discontinuities in the hazard rate in the case of kinks in $H_i(\cdot, \mathbf{x}_i)$ or, in the case of discontinuities in $H_i(\cdot, \mathbf{x}_i)$, positive probability of retirement at the discontinuity date.

The vector (K_1, K_2) is the source of randomness in our econometric model. It is drawn from a joint distribution and its elements are potentially correlated due to, e.g., sorting or other commonalities. It is observed to the agents in the model, but unobserved to the econometrician. As such, it plays the same role as the error in the random utility motivation of the multinomial logit model.

With this setup, the discounted utility for individual i, who transitions to the destination state at t_i, is

$$U^i(t_i, t_j, \mathbf{x}_i, k_i) \equiv \int_0^{t_i} k_i e^{-\rho s} ds + \int_{t_i}^{\infty} H_i(s, \mathbf{x}_i) D(s, t_j) e^{-\rho s} ds$$

where t_j is the time at which the other agent, j, transitions, and (k_1, k_2) is the realization of (K_1, K_2). We implicitly assume that the discount rate ρ and the function H are such that the expression above is well defined. This structure is essentially the same as in Honoré and de Paula (2010). There, it is assumed that the observed outcome, (T_1, T_2), is a Nash equilibrium. That assumption is in the spirit of much of the recent work in industrial organization, but it seems inappropriate in the context of retirement decisions. Given a realization (k_1, k_2) for the random vector (K_1, K_2), we therefore assume that retirement timing is obtained as the solution to the Nash bargaining problem (Nash (1950); see also

Zeuthen (1930)):

$$\max_{t_1,t_2}\left(\int_0^{t_1} k_1 e^{-\rho s}ds \quad + \int_{t_1}^{\infty} H_1(s,\mathbf{x}_i)D(s,t_2)e^{-\rho s}ds - A_1\right)\times \left(\int_0^{t_2} k_2 e^{-\rho s}ds + \int_{t_2}^{\infty} H_2(\cdot,\mathbf{x}_i)D(s,t_1)e^{-\rho s}ds - A_2\right) \tag{1}$$

where A_1 and A_2 are the threat points for spouses 1 and 2, respectively. In the estimation, we set A_i equal to a fraction of the maximum utility individual i would obtain without the increased utility from the externality from the spouse's retirement. This specification of the threat points makes economic sense, and it also saves us from having to deal with the possibility that there are parameter values for which the factors in (1) cannot be made positive. In a more general setting there may be asymmetric bargaining weights that appear as exponents in the objective function. Our analysis could be generalized to include that case, but we ignore this for simplicity and because it is difficult to think about nonparametric features of the data that would allow us to reliably identify such an asymmetry.

The Nash bargaining solution concept is widely used in economics (see, for example, Chiappori, Donni, and Komunjer (2012)). It can be derived from a set of behavioral axioms on the bargaining outcomes (essentially Pareto efficiency, independence of irrelevant alternatives and symmetry) and it is widely adopted in the literature on intra-household bargaining. While it does not pin down a particular negotiation protocol between the parties involved, it can be motivated by the observation that it approximates the equilibrium outcome of a situation where husband and wife make offers to each other in an alternating order and the negotiation breaks down with a certain probability. As this probability goes to zero, the equilibrium converges to the Nash solution (see Binmore, Rubinstein, and Wolinsky (1986)).

One alternative to the Nash bargaining framework used here would be a utilitarian aggregation of the utility functions in the household (i.e., the collective model of Chiappori (1992)). In that case, the retirement dates, (T_1, T_2), would solve:

$$\max_{t_1,t_2} cU^1(t_1,t_2;\mathbf{x}_1,K_1) + U^2(t_2,t_1;\mathbf{x}_2,K_2),$$

where c stands for the relative weight of agent 1's utility. This leads to the following first-order conditions:

$$c \times \frac{\partial U^1(t_1, t_2; \mathbf{x}_1, K_1))}{\partial t_i} + \frac{\partial U^2(t_2, t_1; \mathbf{x}_2, K_2)}{\partial t_i} = 0, \quad i = 1, 2.$$

The setting we propose focuses instead on maximizing $(U^1(t_1, t_2; \mathbf{x}_1, K_1) - A_1) \times (U^2(t_2, t_1; \mathbf{x}_2, K_2) - A_2)$. This leads to the following first-order conditions:

$$\frac{U^2(t_2, t_1; \mathbf{x}_2, K_2) - A_2}{U^1(t_1, t_2; \mathbf{x}_1, K_1) - A_1} \times \frac{\partial U^1(t_1, t_2; \mathbf{x}_1, K_1))}{\partial t_i} + \frac{\partial U^2(t_2, t_1; \mathbf{x}_2, K_2)}{\partial t_i} = 0, \quad i = 1, 2.$$

Consequently, the two are equivalent only if

$$c = \frac{U^2(t_2, t_1; \mathbf{x}_2, K_2) - A_2}{U^1(t_1, t_2; \mathbf{x}_1, K_1) - A_1}.$$

As a result, parameterizing the Nash bargaining approach will impose implicit constraints on c in the corresponding collective model. By the same token, parameterizing the collective model will impose constraints on the corresponding Nash bargaining model. See also Chiappori, Donni, and Komunjer (2012). That paper also establishes identification results when a common set of covariates $\overline{\mathbf{x}}$ affects both the threat points $A_i, i = 1, 2$ and utilities $U^i, i = 1, 2$. Point-identification is achieved using spouse-specific covariates that affect the threat points $A_i, i = 1, 2$, but are excluded from $U^i, i = 1, 2$. In our empirical investigation we rely instead on spouse-specific covariates in $U^i, i = 1, 2$ and no excluded variables in the threat point functions $A_i, i = 1, 2$. Moreover, Chiappori, Donni, and Komunjer (2012) assume that latent variables (i.e., $k_i, i = 1, 2$) are additively separable, which is not our case.

In order to estimate a parameterized version of the Nash bargaining model, we will need to solve it numerically many times. Note that the first term in (1) can be further simplified to

$$\left(K_1 \rho^{-1} \left(1 - e^{-\rho t_1}\right) + \widetilde{H}_1 \left(t_1, \mathbf{x}_1\right) + (\delta - 1) \widetilde{H}_1 \left(\max \left\{t_1, t_2\right\}, \mathbf{x}_1\right) - A_1\right),$$

where $\widetilde{H}_i(t,\mathbf{x}_i) = \int_t^\infty H_i(s,\mathbf{x}_i)e^{-\rho s}ds$ and hence[5] $\widetilde{H}_i'(t,\mathbf{x}_i) = -H_i(t,\mathbf{x}_i)e^{-\rho t}$. An analogous simplification applies to the second term. In other words, the objective function is given by

$$N(t_1,t_2) = \overbrace{\left(K_1\rho^{-1}\left(1-e^{-\rho t_1}\right) + \widetilde{H}_1(t_1,\mathbf{x}_i) + (\delta-1)\widetilde{H}_1(\max\{t_1,t_2\},\mathbf{x}_1) - A_1\right)}^{\equiv I} \times \underbrace{\left(K_2\rho^{-1}\left(1-e^{-\rho t_2}\right) + \widetilde{H}_2(t_2,\mathbf{x}_2) + (\delta-1)\widetilde{H}_2(\max\{t_1,t_2\},\mathbf{x}_2) - A_2\right)}_{\equiv II}.$$

If the two spouses retire sequentially, say, $T_1 < T_2$, the first-order condition with respect to t_1 is

$$\left(K_1e^{-\rho t_1} - H_1(t_1,\mathbf{x}_1)e^{-\rho t_1}\right)\left(\int_0^{T_2} K_2e^{-\rho s}ds + \int_{T_2}^\infty H_2(s,\mathbf{x}_2)\delta e^{-\rho s}ds - A_2\right) = 0.$$

This implies that either

$$K_1 = H_1(T_1,\mathbf{x}_1)$$

or

$$\int_0^{T_2} K_2e^{-\rho s}ds + \int_{T_2}^\infty H_2(s,\mathbf{x}_2)\delta e^{-\rho s}ds = A_2.$$

The second possibility is ruled out since we specify the threat points so that each person gets a higher utility than his or her threat point at the Nash bargaining solution. The first-order condition with respect to t_2 gives

$$H_1(t_2,\mathbf{x}_1)e^{-\rho t_2}(1-\delta) \times (II) + (I) \times \left(K_2e^{-\rho t_2} - H_2(t_2,\mathbf{x}_2)\delta e^{-\rho t_2}\right) = 0. \tag{2}$$

The t_2 that solves this equation is smaller than the value obtained in Honoré and de Paula (2010): $H_2^{-1}(K_2/\delta,\mathbf{x}_2)$.[6] Intuitively, the reason is that with Nash bargaining, the second spouse to retire is willing to forgo some utility if the increase in utility to the other spouse is sufficiently high. Mathematically, we see this by noting that $H_1(t_2,\mathbf{x}_1) > 0$,

[5]When the covariates are time-varying, $\widetilde{H}_i'(t,\mathbf{x}_i)$ denotes the total derivative of $\widetilde{H}_i(t,\mathbf{x}_i)$.

[6]In Honoré and de Paula (2010), we only consider the case where $H_i(t,\mathbf{x}_i) = Z(t)\varphi(\mathbf{x}_1)$.

$e^{-\rho t_2} > 0$, and $1 - \delta < 0$. Moreover, II must be positive in equilibrium. This implies that $H_1(t_2, \mathbf{x}_1)e^{-\rho t_2}(1-\delta) \times (II) \leq 0$ at the solution. So for the first-order condition to be zero, the product $(I) \times (K_2 e^{-\rho t_2} - H_2(t_2, \mathbf{x}_2)\delta e^{-\rho t_2})$ should be positive. Since I and $e^{-\rho t_2}$ are both positive, K_2 therefore must be greater than $H_2(t_2, \mathbf{x}_2)\delta$. Or equivalently, $T_2 < H_2^{-1}(K_2/\delta, \mathbf{x}_2)$. This implies that

$$\begin{aligned} T_1 &= H_1^{-1}(K_1, \mathbf{x}_1) \\ T_2 &\leq H_2^{-1}(K_2/\delta, \mathbf{x}_2), \end{aligned}$$

which gives the same timing choice for the first retiree as in Honoré and de Paula (2010) but an earlier one for the spouse. A similar set of calculations is obtained for $T_2 < T_1$.[7]

A third possibility is for the spouses to retire jointly. In this case,

$$\begin{aligned} T &= \arg\max_t N(t,t) \\ &= \arg\max_t \left(K_1\rho^{-1}\left(1-e^{-\rho t}\right) + \delta\widetilde{H}_1(t, \mathbf{x}_1) - A_1\right)\left(K_2\rho^{-1}\left(1-e^{-\rho t}\right) + \delta\widetilde{H}_2(t, \mathbf{x}_2) - A_2\right). \end{aligned}$$

The derivative of this with respect to t is

$$\begin{aligned} &e^{-\rho t}\left(K_1 - \delta H_1(t, \mathbf{x}_1)\right)\left(K_2\rho^{-1}\left(1-e^{-\rho t}\right) + \delta\widetilde{H}_2(t, \mathbf{x}_2) - A_2\right) \\ &+e^{-\rho t}\left(K_1\rho^{-1}\left(1-e^{-\rho t}\right) + \delta\widetilde{H}_1(t, \mathbf{x}_1) - A_1\right)\left(K_2 - \delta H_2(t, \mathbf{x}_2)\right). \end{aligned}$$

When $t < H_1^{-1}(K_1/\delta, \mathbf{x}_1)$ and $t < H_2^{-1}(K_2/\delta, \mathbf{x}_2)$ this derivative is positive, and when $t > H_1^{-1}(K_1/\delta, \mathbf{x}_1)$ and $t > H_2^{-1}(K_2/\delta, \mathbf{x}_2)$, it is negative. The optimum is therefore in the

[7] For computation purposes we also notice that the objective function is unimodal on t_2. If we start at the critical value, increasing t_2 reduces the function. This is because, for small ρ, $H_1(t_2, \mathbf{x}_1)e^{-\rho t_2}(1-\delta)$ becomes *more* negative and II becomes *more* positive, so the product becomes *more* negative. For the second term, I decreases and $k_2 e^{-\rho t_2} - H_2(t_2, \mathbf{x}_2)\delta e^{-\rho t_2}$, *which is positive*, decreases. Their product then decreases. Consequently, the derivative, which is the sum of these two products, becomes negative, and the objective function is decreasing. Analogously, we can also determine that the objective function is increasing for values below the critical value.

interval

$$\min\left\{H_1^{-1}\left(K_1/\delta,\mathbf{x}_1\right),H_2^{-1}\left(K_2/\delta,\mathbf{x}_2\right)\right\}\leq t\leq\max\left\{H_1^{-1}\left(K_1/\delta,\mathbf{x}_1\right),H_2^{-1}\left(K_2/\delta,\mathbf{x}_2\right)\right\}.$$

This is useful in the numerical solution to the optimal solution condition on joint retirement.

Figure 2 illustrates these cases and plots the optimal transition times, T_1 and T_2, as a function of K_2 as K_1 is held fixed. For low values of K_2, $T_1 > T_2$: in the retirement example, labor force attachment is higher for spouse 1 than for spouse 2. When K_1 is large, on the other hand, $T_1 < T_2$ and spouse 1 retires sooner. For intermediary values of K_1, $T_1 = T_2$ and the two spouses retire at the same time. This generates probability distributions such as those in Figure 3. Unconditionally, the probability density function for T_1 is smooth. Conditionally on $T_2 = t_2$, though, a point mass at $T_1 = t_2$ arises.

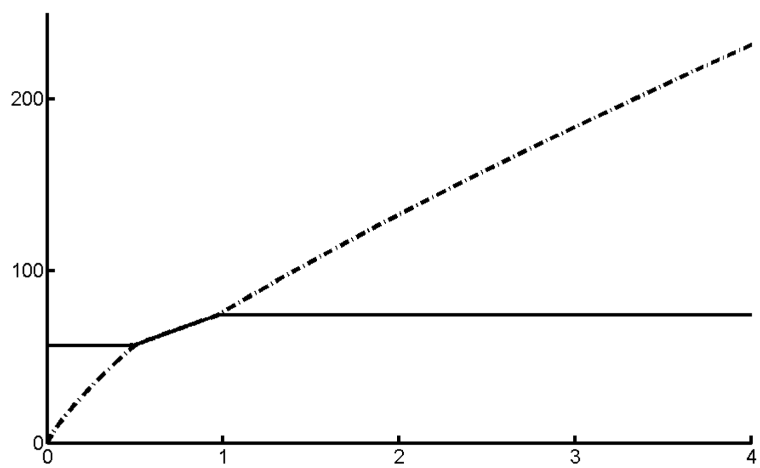

Figure 2: T_1 (solid line) and T_2 (dashed line) as Functions of K_2 (For K_1 Fixed)

The set of realizations of (K_1, K_2) for which $T = T_1 = T_2$ is an optimum is larger than the set obtained in the non-cooperative setup from Honoré and de Paula (2010). This is illustrated in Figure 4, where the area between the dotted lines is the joint retirement region in Honoré and de Paula (2010) and the area between solid lines is the joint retirement region in the current paper. Also, in that paper any date within a range $[\underline{T} < \overline{T}]$ was sustained as an equilibrium for pairs (K_1, K_2) inducing joint retirement. In contrast, the equilibrium

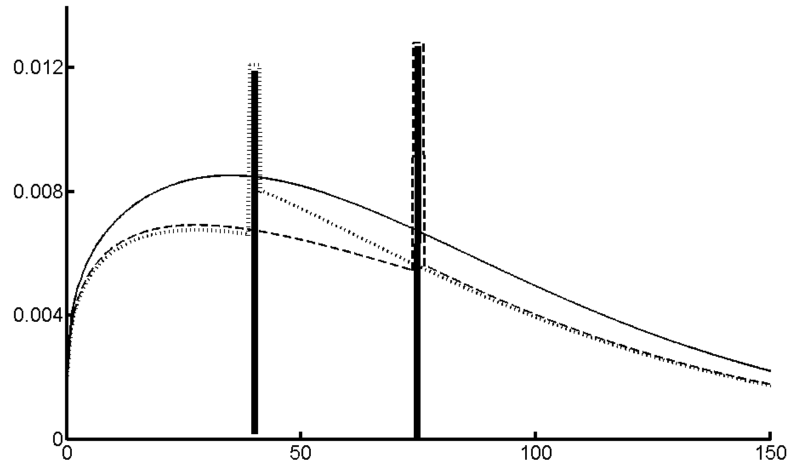

Figure 3: Marginal Density for T_1 (solid line) and Conditional Given $T_2 = 45$ (dotted line) and $T_2 = 75$ (dashed line).

joint retirement date for a given realization of (K_1, K_2) is uniquely pinned down in the setup here. Because Nash bargaining implies Pareto efficiency and because $\underline{T}$ is the Pareto dominant outcome among the possible multiple equilibria in the game analyzed by Honoré and de Paula (2010), it should be the case that joint retirement in the Nash bargaining model occurs on or before $\underline{T}$. In comparison to the non-cooperative paradigm adopted in our previous paper, Nash bargaining allows spouses to "negotiate" an earlier retirement date, which is advantageous to the household.

Finally, we note that when $H_i(t, \mathbf{x}_i) = Z_i(t)\, \varphi_i(\mathbf{x}_1)$ and $\delta = 1$, the optimal retirement dates will correspond to

$$\log Z_i(T_i) = -\log \varphi_i + \log K_i, \quad i = 1, 2.$$

K_i following a unit exponential distribution gives a proportional hazard model. For a general distribution of K_i, this yields the generalized accelerated failure time model of Ridder (1990). This is the sense in which the approach discussed in this section can be thought of as a simultaneous equations version of a generalized accelerated failure time model.

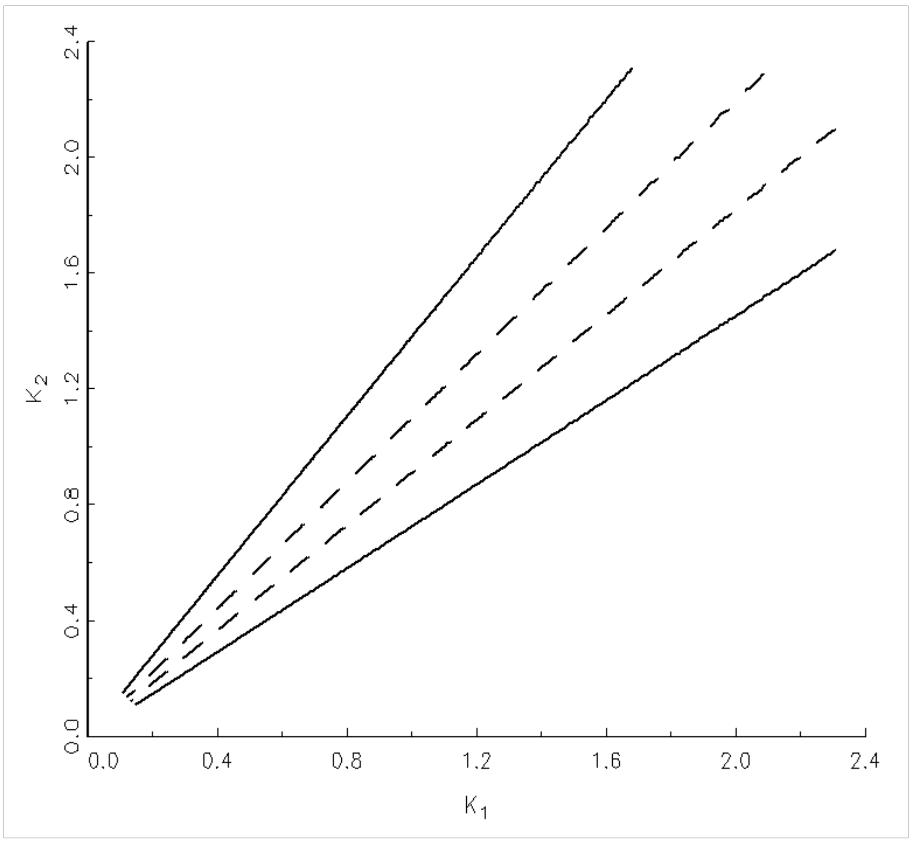

Figure 4: Joint Retirement Region. This paper (solid line) and Honoré and de Paula (2010) (dashed line).

2.2 Parameterization

In the construction of an econometric model for multiple durations that start at different times, one must decide whether to measure time in terms a common calendar time or in terms of the individual durations. Since the motivation of this paper is that events sometimes happen at the same time, it is typically most convenient to measure time in terms of calendar time. In our empirical analysis, we measure time in terms of "family age," which is set to zero when the older partner in the couple reaches age 60. We then keep track of the age of the other spouse by using the age difference between the husband and the wife as a covariate. Alternatively, we could have worked with the individuals' ages, but that would have been more cumbersome because the economic motivation is that the person may enjoy utility from being retired at the same time as his or her spouse, and not from being retired at the same age. Throughout, we use $i = 1, 2$ to denote the two spouses in a married couple. n is used to index couples.

In the empirical application, we specify $H_i(t, \mathbf{x}_i)$ as $Z_i(t)\varphi_i(\mathbf{x}_1)$, which yields the link to the generalized accelerated failure time model. If we further parameterize $Z_i(t)$ as $Z(t; \theta_{1i}) = t^{\theta_{1i}}$ and if $\delta = 1$ then the model developed in the previous section will deliver the simple Weibull regression model with integrated baseline hazards $t^{\theta_{1i}}$ for the two durations as a special case when $K_i \sim \exp(1)$. Our parameterization therefore takes those as the point of departure.

The parameterization $Z(t; \theta_{1i}) = t^{\theta_{1i}}$ will also yield a convenient expression for $\widetilde{Z}_i$:

$$\widetilde{Z}_i(t) = \int_t^\infty s^{\theta_{1i}} e^{-\rho s} ds = \left(\frac{1}{\rho}\right)^{\theta_{1i}+1} \Gamma(\theta_{1i} + 1, \rho t),$$

where the upper incomplete gamma function is defined by $\Gamma(\theta_{1i}, x) = \int_x^\infty s^{\theta_{1i}-1} e^{-s} ds$.[8]

As discussed later, the structure of the US Social Security system introduces incen-

[8]This expression can be further manipulated by noting that if the random variable X is Gamma distributed with parameters α and $\beta = 1$

$$\overline{F}_{\Gamma(\alpha,1)}(x) = P(X > x) = \frac{1}{\Gamma(\alpha)} \int_x^\infty s^{\alpha-1} e^{-s} ds = \frac{\Gamma(\alpha, x)}{\Gamma(\alpha)}.$$

tives to retire at certain ages. This could be accounted for by introducing a time-varying dummy variable as one of the explanatory variables or by allowing for jumps in Z. Since the other explanatory variables in the application are time-invariant, we find it notationally more convenient to incorporate a discontinuity in Z_i. In the empirical application below we therefore introduce a jump in Z at τ, the time (measured in family calendar months) at which the individual turns 62. One way to do this would be to augment Z as

$$Z\left(t;\left(\alpha,\gamma\right),\tau\right)=Z_1\left(t;\alpha\right)+\gamma 1\left\{t\geq\tau\right\},$$

where $\gamma\geq 0$ is a parameter to be estimated. We choose a slightly different version in which $1\left\{t\geq\tau\right\}$ is replaced by a smooth function that increases from 0 to 1 over the interval τ to $\tau+1$,

$$Z\left(t;\left(\alpha,\gamma\right),\tau\right)=Z_1\left(t;\alpha\right)+Z_2\left(t;\gamma,\tau\right)=t^a+\gamma F(t;\tau)$$

and

$$F(t;\tau)=\begin{cases}0 & \text{for} \quad t<\tau\\ 2\left(t-\tau\right)^2 & \text{for} \quad \tau<t<\tau+1/2\\ 1-2\left(t-1-\tau\right)^2 & \text{for} \quad \tau+1/2<t<\tau+1\\ 1 & \text{for} \quad \tau+1<t\end{cases}.$$

As discussed later, the dataset will deliver durations rounded to a month. Our choice of F is therefore observationally equivalent to the step function, $1\left\{t\geq\tau\right\}$, but it makes it easy to calculate Z^{-1} numerically.

Consequently,

$$\widetilde{Z}\left(t;\alpha\right)=\left(\frac{1}{\rho}\right)^{\alpha+1}\Gamma\left(\alpha+1,\rho t\right)=\left(\frac{1}{\rho}\right)^{\alpha+1}\Gamma\left(\alpha+1\right)\overline{F}_{\Gamma(\alpha+1,1)}\left(\rho t\right)$$

which is useful since both $\Gamma\left(\cdot\right)$ and $\overline{F}_{\Gamma(\cdot,1)}\left(\cdot\right)$ are preprogrammed in many software packages.

To calculate $\widetilde{Z}_2(t;\gamma,\tau)$ we note that for $t<\tau$,

$$\widetilde{Z}_2(t;\gamma,\tau)=\int_{\tau}^{\tau+1/2}\left(2s^2-4\tau s+2\tau^2\right)e^{-\rho s}ds \\ +\int_{\tau+1/2}^{\tau+1}\left(-2s^2+4(\tau+1)s+1-2(1+\tau)^2\right)e^{-\rho s}ds+\int_{\tau+1}^{\infty}e^{-\rho s}ds$$

for $\tau<t<\tau+1/2$,

$$\widetilde{Z}_2(t;\gamma,\tau)=\int_{t}^{\tau+1/2}\left(2s^2-4\tau s+2\tau^2\right)e^{-\rho s}ds \\ +\int_{\tau+1/2}^{\tau+1}\left(-2s^2+4(\tau+1)s+1-2(1+\tau)^2\right)e^{-\rho s}ds+\int_{\tau+1}^{\infty}e^{-\rho s}ds$$

for $\tau+1/2<t<\tau+1$,

$$\widetilde{Z}_2(t;\gamma,\tau)=\int_{t}^{\tau+1}\left(-2s^2+4\tau s+1-2(1-\tau)^2\right)e^{-\rho s}ds+\int_{\tau+1}^{\infty}e^{-\rho s}ds$$

and finally for $\tau+1<t$

$$\widetilde{Z}_2(t;\gamma,\tau)=\int_{t}^{\infty}e^{-\rho s}ds.$$

All the integrals have the form $\int s^j e^{-\rho s}ds$ where j is an integer. Hence they can all be expressed in closed form.

To allow for positive correlation between the unobserved variables K_1 and K_2 (induced, e.g., by sorting), we use a Clayton-Cuzick copula function (see Clayton and Cuzick (1985)). More precisely, we model the joint survivor distribution function of K_1 and K_2 as:

$$F_{K_1,K_2}(k_1,k_2;\tau)=K(\exp(-k_1),\exp(-k_2);\tau),$$

where

$$K(u,v;\tau)=\begin{cases}\left(u^{-\tau}+v^{-\tau}-1\right)^{-1/\tau} & \text{for} \quad \tau>0\\ uv & \text{for} \quad \tau=0.\end{cases} \tag{3}$$

The unobservables are independent when $\tau=0$. When $\tau>0$, there is positive dependence between variables K_1 and K_2. Specifically, Kendall's rank correlation for the Clayton-Cuzick copula is equal to $\tau/(2+\tau)$ (see, for example, Trivedi and Zimmer (2006)). This copula is commonly used to introduce dependence in the duration literature. Finally, we take $\varphi_i(x_i) = \exp(\theta_{2i}^{\top} x_i)$. This implies that when $\delta = 1$ and $\tau = 0$, the durations follow simple independent proportional hazard Weibull models (Lancaster (1990), p.44). This is the sense in which our approach generalizes simple standard econometric duration models. We also note that, even if $\delta = 1$, the copula used here plays the dual role of introducing correlation between the unobserved variables K_1 and K_2 and allowing for unobserved heterogeneity in the hazard rates to retirement.

Clayton and Cuzick (1985) motivate (3) as the unique copula with a certain constant odds ratio. However, they also point out that it is consistent with a model in which the dependence between the two durations is driven by common unobserved heterogeneity (with a specific distribution). It is therefore tempting to ask whether it is feasible to introduce additional unobserved heterogeneity. There are two reasons why we do not consider this possibility in our empirical application. The first is that since our model includes the mixed Weibull model as a special case, a nonparametric specification of the heterogeneity distribution will make root–n consistent estimation of the model parameters infeasible (see Hahn (1994)). Since some of the parameters of the model are already imprecisely estimated, this suggests that a flexible parametric specification of the heterogeneity distribution would not be fruitful. The second, and related, reason is that we estimate our model using indirect inference. To estimate a model with unobserved heterogeneity, we would therefore have to specify an auxiliary model whose parameters are informative about the heterogeneity distribution. As mentioned earlier, when $\delta = 1$ the optimal retirement dates will correspond to

$$\log Z_i(t_i) = -\log \varphi_i + \log K_i, \quad i = 1, 2.$$

If the jump in the baseline hazard (γ) is zero, then our parameterization yields

$$\log(t_i) = -\frac{x_i'\theta_{2i}}{\theta_{1i}} + \frac{\log K_i}{\theta_{1i}}, \quad i = 1, 2,$$

where $\log K_i$ distributed according to minus an extreme value distribution. On the other hand, a model with additional unobserved heterogeneity would have

$$\log(t_i) = -\frac{x_i'\theta_{2i}}{\theta_{1i}} + \frac{\log K_i}{\theta_{1i}} + v_i, \quad i = 1, 2. \tag{4}$$

In other words, the distribution of the unobserved heterogeneity would be identified from deviations of the distribution of the error term in (4) from an extreme value distribution. Given the heavy censoring (more than 50%; see Section 3), this does not seem fruitful.

2.3 Estimation: Indirect Inference

Because the likelihood function for the model developed in the previous section is not easily computed in closed form, we resort to simulation-assisted methods. One potential strategy would be to use simulated maximum likelihood (SML), where one non-parametrically estimates the conditional likelihood via kernel methods applied to simulations of T_1 and T_2 at particular parameter values and searches for the parameter value that maximizes the (simulated) likelihood. We opt for a different strategy for two main reasons. First, our likelihood displays some non-standard features. For example, there is a positive probability for the event $\{T_1 = T_2\}$. Second, consistency of the SML estimator requires a large number of simulations, which can be computationally expensive.

To estimate our model we therefore employ an indirect inference strategy (see Gourieroux, Monfort, and Renault (1993); Smith (1993); and Gallant and Tauchen (1996)). Rather than estimating the maximum likelihood estimator for the true model characterized by parameter θ, one estimates an approximate (*auxiliary*) model with parameter β. Let $n = 1, \ldots, N$ index a sample of households (couples). Then, under the usual regularity conditions,

$$\widehat{\beta} = \arg\max_b \sum_{n=1}^{N} \log \mathcal{L}_a \left(b; z_n\right) \xrightarrow{p} \arg\max_b E_{\theta_0} \left[\log \mathcal{L}_a \left(b; z_n\right)\right] \equiv \beta_0 \left(\theta_0\right) \tag{5}$$

where $\mathcal{L}_a$ is a pseudo–likelihood function (parameterized by b) for the auxiliary model, z_n is the data for observation n, and the expectation E_{θ_0} is taken with respect to the true model. $\beta_0\left(\theta_0\right)$ is known as the pseudo–true value and the key is that it depends on the true parameters of the data–generation process (θ_0). The basic idea, then, is that if one knew the pseudo–true value as a function of θ_0, it could be used to solve the equation

$$\widehat{\beta} = \beta_0\left(\widehat{\theta}\right)$$

and obtain an estimator for θ_0. In our case, we do not know $\beta_0(\theta)$, but we can easily approximate this function using simulations from the structural model. For a particular value of the parameters of the structural model, θ, we generate R draws

$$\left\{\left(z_{1r}\left(\theta\right), z_{2r}\left(\theta\right), \ldots, z_{Nr}\left(\theta\right)\right)\right\}_{r=1}^{R}$$

from our structural model. In practice this is done by transforming uniform random variables. These are then kept fixed as one varies θ. The parameter, θ, enters through the transformation of these uniform random variables. We can then estimate the function

$$\beta_0(\theta) \equiv \arg\max_b E_{\theta}\left[\log \mathcal{L}_a\left(b; z_n\right)\right]$$

by

$$\widetilde{\beta}_R\left(\theta\right) = \arg\max_b \frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{n=1}^{N}\left(\log \mathcal{L}_a\left(b; z_{nr}\left(\theta\right)\right)\right).$$

This suggests finding $\widehat{\theta}$ such that the generated data set using $\widehat{\theta}$ gives the same estimate in the auxiliary model as we got in the real sample, $\widehat{\beta} = \widetilde{\beta}_R\left(\widehat{\theta}\right)$. When the dimensionality of β is greater than the dimension of θ, this is not possible, and one then estimates θ by a minimum distance approach that makes the difference between $\widehat{\beta}$ and $\widetilde{\beta}_R\left(\theta\right)$ as small as possible.

While this approach is conceptually straightforward, it requires one to estimate β for each potential value of θ. This can be computationally burdensome and we therefore adopt a slightly different version based on the first-order conditions from estimating the auxiliary model. The expression (5) implies that

$$\frac{1}{N}\sum_{n=1}^{N}\mathcal{S}_a\left(\widehat{\beta};z_n\right)=0,$$

and that $\widehat{\beta}$ converges to the solution to $E_\theta\left[\mathcal{S}_a\left(b;z_n\right)\right]=0$, where S_a is the pseudo–score associated with $\mathcal{L}_a$. Of course, the solution $E_\theta\left[\mathcal{S}_a\left(b;z_n\right)\right]=0$ is just $\beta_0\left(\theta_0\right)$ from equation (5). As before, we estimate $E_\theta\left[\mathcal{S}_a\left(\cdot;z_n\right)\right]$ as a function of θ using

$$\frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{n=1}^{N}\mathcal{S}_a\left(\cdot;z_{nr}\left(\theta\right)\right)$$

and θ_0 is estimated by making it as close to zero as possible. Specifically, if $\dim\left(\mathcal{S}_a\right)>\dim\left(\beta\right)$, we minimize

$$\left(\frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{n=1}^{N}\mathcal{S}_a\left(\widehat{\beta};z_{nr}\left(\theta\right)\right)\right)^{\top}W\left(\frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{n=1}^{N}\mathcal{S}_a\left(\widehat{\beta};z_{nr}\left(\theta\right)\right)\right) \qquad (6)$$

over θ. The weighting matrix W is a positive definite matrix playing the usual role in terms of estimator efficiency. The optimal W can be calculated using the actual data (before estimating θ) and the asymptotic properties follow from standard GMM arguments (see Gourieroux and Monfort (1996) for details). This strategy is useful because we only estimate the auxiliary model once using the real data. After that, we evaluate its first-order condition using simulated data from the structural model for different values of θ.

The retirement times used in the empirical application are interval censored. When doing the indirect inference, we mimic this by evaluating (6) at interval censored simulated durations. Finally, the outcome variable in our empirical analysis is censored. To use simulation-based inference we must be able to simulate data that have been censored by the same process. In practice this means that we must either model the censoring process parametrically or observe the censoring times even for those observations that are uncensored in the data. As discussed below, our application falls into the second category.

2.3.1 Auxiliary Model

Our auxiliary model is composed of four reduced-form models that are chosen to capture the features of the data that are our main concern: the duration until retirement for each of the two spouses, the idea that some married couples choose to retire jointly, and finally the idea that the retirement durations may exhibit correlation (conditionally on the covariates) even when they are not equal. For the first two, we use a standard proportional hazard model for each spouse with a Weibull baseline hazard and the usual specification for the covariate function. For the third, we use an ordered logit model as suggested by our paper Honoré and de Paula (2010). For the fourth feature, we exploit the covariance in the residuals in regressions of the two retirement durations on all the covariates of the model. We present the models in detail below.

2.3.2 Weibull Proportional Hazard Model

For each spouse i, the hazard for retirement conditional on x_i is assumed to be $\lambda_i\left(t|\, x_i\right) = \alpha_i t^{\alpha_i - 1}\exp\left(x_i'\beta_i\right)$. The (log) density of retirement for spouse i conditional on x_i, $\log f_i\left(t|x_i\right)$, is then given by:

$$\log\left\{\lambda_i\left(t\right)\exp\left(x_i'\beta_i\right)\exp\left(-Z_i\left(t\right)\exp\left(x_i'\beta_i\right)\right)\right\} = \log\alpha_i + \left(\alpha_i - 1\right)\log t + x_i'\beta_i - t^{\alpha_i}\exp\left(x_i'\beta_i\right)$$

The (conditional) survivor function can be obtained analogously, and it is given by:

$$\log S_i\left(t|x_i\right) = \log\left\{\exp\left(-Z_i\left(t\right)\exp\left(x_i'\beta_i\right)\right)\right\} = -t^{\alpha_i}\exp\left(x_i'\beta_i\right)$$

Letting $c_{i,n} = 1$ if the observed retirement date for spouse i in household n is (right-)censored, and $= 0$ otherwise, we obtain the log-likelihood function:

$$\log \mathcal{L} = \sum_{n=1}^{N} (1 - c_{i,n}) \left(\log \alpha_i + (\alpha_i - 1) \log (t_{i,n}) + x'_{i,n}\beta_i \right) - \sum_{n=1}^{N} t_{i,n}^{\alpha_i} \exp \left(x'_{i,n}\beta_i \right)$$

First- and second-order derivatives used in the computation of the MLE for this auxiliary model are presented in the Appendix.

2.3.3 Ordered Logit Model Pseudo MLE

In the spirit of the estimation strategy suggested in Honoré and de Paula (2010), we also use an ordered logit model as an auxiliary model. Whereas the Weibull model will convey information on the timing of retirement, this second auxiliary model will provide information on the pervasiveness of joint retirement and help identify the taste interactions leading to this phenomenon (i.e., δ). Define

$$y_n = \begin{cases} 1, & \text{if } t_1 > t_2 + 1 \\ 2, & \text{if } |t_1 - t_2| \leq 1 \\ 3, & \text{if } t_2 > t_1 + 1 \end{cases}$$

Incorrectly assuming an ordered logit model for y_n yields

$$P(y_n = 1 \text{ or } y_n = 2) = \Lambda \left(x'_n \gamma_1 \right) \quad \text{and} \quad P(y_n = 2) = \Lambda \left(x'_n \gamma_1 - \gamma_0 \right)$$

where $\Lambda(\cdot)$ is the cumulative distribution function for the logistic distribution.

This allows us to construct the following pseudo-likelihood function:

$$\mathcal{Q}(\gamma) = \sum_{y_n=0} \log \left(1 - \Lambda \left(x'_{0n}\gamma \right) \right) + \sum_{y_n \neq 0} \log \left(\Lambda \left(x'_{0n}\gamma \right) \right) + \sum_{y_n \neq 2} \log \left(1 - \Lambda \left(x'_{1n}\gamma \right) \right) + \sum_{y_n=2} \log \left(\Lambda \left(x'_{1n}\gamma \right) \right)$$

where

$$x_{0n} = \left(x_n' \vdots \mathbf{0}\right)' \qquad x_{1n} = \left(x_n' \vdots \mathbf{1}\right)' \qquad \gamma = \left(\gamma_1' \vdots - \gamma_0\right)'$$

As before, first- and second-order derivatives are presented in the Appendix.

The explanatory variables in the different parts of the auxiliary model need not be the same, and they need not coincide with the explanatory variables in the model to be estimated. In the empirical section below, the covariates in the Weibull auxiliary models are each spouses's own values of the explanatory variables in the model of interest. We use a constant only as an explanatory variable in the ordered logit model. This leaves the number of overidentifying restrictions constant across specifications.

In the data and in the simulations, y is defined using the failure time (i.e., the minimum between censoring and retirement dates). Censored observations do not pose problems when the other person in the household is uncensored and retires earlier, since in that case we can determine that retirement happened sequentially. Whereas we can always mark whether retirement was sequential or simultaneous in the simulations, when censoring happens before the retirement of the uncensored partner or both are censored, we cannot determine in the data whether retirement was sequential. Since we use the failure time in both the data and the simulations, censoring introduces the same degree of "noise" in the definition of y in the data and in the simulations.

2.3.4 Covariance in Failure Times

To allow for correlation in the unobservable variables K_1 and K_2, we use copula functions. We augment our auxiliary models with the covariance in failure times (including censored observations in both the data and the simulation moments) to perform the estimation. Specifically, we match the covariance between the residuals from a regression of (censored log) failure time on all covariates for husband and wife. An alternative is to use the residuals from regressions on spouse-specific variables and/or to define generalized residuals from a proportional hazard model estimated by maximum likelihood. The reason why we did not choose those approaches is that the asymptotic distribution for the covariance would then depend on nuisance parameters (i.e., the regression coefficients). This is not the case if we use the same set of covariates for husband and wife and estimate the model by OLS. Our procedure is therefore asymptotically equivalent to matching the true errors from those regressions (projections).

In the notation of an objective function for an auxiliary model, we maximize

$$\mathcal{C}(\rho) = -\sum_{n} \left(\hat{u}_{1,n}\hat{u}_{2,n} - \rho\right)^2$$

where $\hat{u}_{i,n} = \ln(t_{i,n}) - x_n^\top (\sum_h x_h x_h^\top)^{-1} (\sum_h x_h \ln(t_{i,h}))$ for $i = 1, 2$ with $t_{i,n}$ representing the failure time (earliest between retirement and censoring time) for partner i in couple n, x_n representing the covariates for couple n and $\hat{u}_{i,nr}$ is defined analogously on the simulated observations.

2.3.5 Failure Probability at Early Retirement Age

In the United States, individuals can claim Social Security benefits as soon as they turn 62 years old. Whereas this implies a penalty vis-à-vis the official retirement age,[9] it is noticeable that many individuals elect to retire as soon as they reach 62 years of age. (In our data, this is visible from the steep increase in the CDF for the retirement year of husbands in Figure 6.) To accommodate this possibility, we allow for a discontinuity in $Z(\cdot)$ at the early retirement age. To capture this feature of the model, we employ the probability of retirement in the (closed) interval between one month before and six months after turning 62. In the notation of an objective function for an auxiliary model, we maximize

$$\mathcal{S}(\psi) = -\sum_{n} \sum_{i=1}^{2} \left(1\left\{age_{i,n}^{62} - 1 \le t_{i,n} \le age_{i,n}^{62} + 6\right\} - \psi_i\right)^2$$

[9]The official retirement age was 65 years old for individuals born in 1937 or earlier, and for persons born after that year, it gradually increases to 67 years old, which is the retirement age for those born in 1960 or after.

where $age_{i,n}^{62}$ is the age (in months and measured in family-time) at which individual i in family n turns 62.

2.3.6 Overall Auxiliary Model

The overall auxiliary model objective function is then defined by the pseudo-log-likelihood function

$$\log \mathcal{L}_{men}(\alpha_1, \beta_1) + \log \mathcal{L}_{women}(\alpha_2, \beta_2) + \mathcal{Q}(\gamma) + \mathcal{C}(\rho) + \mathcal{S}(\psi)$$

and the moment conditions used for estimating the parameters of the structural model are the first-order conditions for maximizing this.

As indicated above, we choose as our weighting matrix $W = \widehat{J}_0^{-1}$, where

$$\widehat{J}_0 = \widehat{V}\left[\begin{pmatrix} \frac{\partial \log \mathcal{L}_{mn}}{\partial(\alpha_1,\beta_1)} \\ \frac{\partial \log \mathcal{L}_{wn}}{\partial(\alpha_2,\beta_2)} \\ \frac{\partial \mathcal{Q}_n}{\partial \gamma} \\ \frac{\partial \mathcal{C}_n}{\partial \rho} \\ \frac{\partial \mathcal{S}_n}{\partial \psi} \end{pmatrix}\right]$$

The (asymptotic) standard errors of the structural estimates are calculated using the formulae in Gourieroux and Monfort (1996).

The computational details are described in the appendix.

3 Data

We estimate the model using eight waves of the Health and Retirement Study (every two years from 1992 to 2006) and keep households where at least one individual was 60 years old or more. Retirement is observed at a monthly frequency. In accordance with this sampling scheme, we aggregate the simulated retirement dates to the month level before evaluating (6). We use the retirement classification suggested by the Rand Corporation. This classifies a respondent as retired if she/he is not working and not looking for work or there is any mention of retirement through the employment status or the questions that ask the respondent whether he or she considers him- or herself to be retired.[10] To avoid left-censoring, selected households also had both partners in the labor force at the initial period. Right-censoring occurs when someone dies or is not retired at his or her last interview before the end of the survey. We excluded individuals who were part of the military. Finally, we exclude households with multiple couples and individuals with multiple spouses during the period of analysis, couples with conflicting information over marital status or other joint variables, and couples of the same gender. This leaves us with 1,284 couples. Of those, 407 couples have both the husband's and the wife's uncensored retirement dates. Among the uncensored couples, 31 couples ($\approx 7.6\%$) retire jointly.[11] Figure 5 plots the retirement month of the husbands against the retirement month of the wives for those couples whose retirement month is uncensored for both spouses (January 1931 is month 1). The points along the 45-degree line are the joint retirements.

We measure covariates in the first "household year": when the older partner reaches the age of 60.[12] The covariates we use are:

1. the age difference in the couple (husband's age minus wife's age in years);
2. dummies for race (non-Hispanic black, Hispanic and other race with non-Hispanic whites as the omitted category);
3. dummies for education (high school or GED, some college and college or above with less than high school as the omitted category);
4. indicators of region (NE, SO, and WE with MW or other region as the omitted category);

[10] Specifically, we use the classification provided by the variable `RwLBRF`.

[11] There are 451 additional couples with only one censored spouse. If those are presumed to have retired sequentially, the proportion of joint retirements among couples with at most one censored spouse is 3.6%. Taking into account the remaining households where both individual retirement dates are censored would place the proportion of simultaneous retirements somewhere between 2.4% (if all additional households are assumed to retire sequentially) and 13.8% (if all additional households are assumed to retire simultaneously).

[12] We take the measurements from the first interview after the older spouse turns 60.

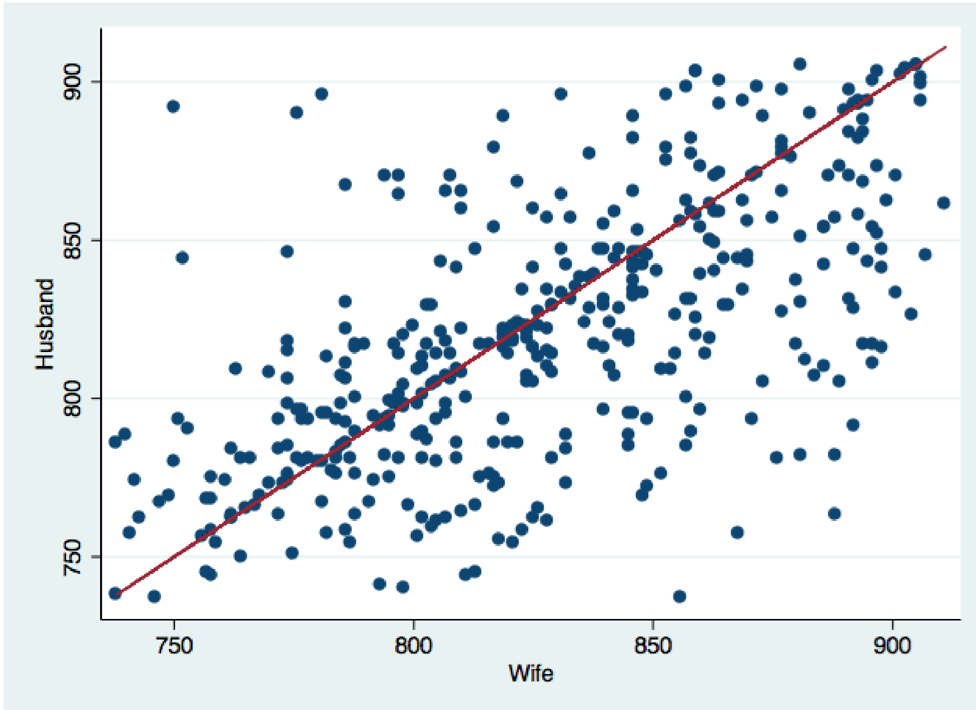

Figure 5: Retirement Months: Husband vs Wife

5. self-reported health dummies (good health, very good health, with poor health as the omitted category);

6. an indicator for whether the person has health insurance;

7. the total health expenditure per individual in the previous 12 months for the first two waves and the previous 2 years for the subsequent years[13] (inflation adjusted using the CPI to Jan/2000 dollars);

8. indicators for whether the person had a defined contribution (DC) or defined benefit (DB) plan; and

9. financial wealth (inflation adjusted using the CPI to Jan/2000 dollars).[14] This measure

[13] We use the transformation sgn(total health expenditure)$\times\sqrt{\text{abs(total health expenditure)}}$. This transformation is in the spirit of a logarithmic transformation of positive variables and implies that large quantities have a decreasing effect. In contrast to a log transformation, it allows us to handle zeroes. In the computations, we also divide the transformed variable by 10^2 to avoid overflow.

[14] For financial wealth we use the transformation sgn(financial wealth)$\times\sqrt{\text{abs(financial wealth)}}$. This transformation is in the spirit of a logarithmic transformation of positive variables and implies that large quantities have a decreasing effect. In contrast to a log transformation, it allows us to handle negative

includes the value of checking and savings accounts, stocks, mutual funds, investment trusts, CDs, government bonds, Treasury bills and all other savings minus the value of debts such as credit card balances, life insurance policy loans or loans from relatives. It does not include housing wealth or private pension holdings.

In Section 4, we also examine whether our focus on covariates recorded at the initial household year might affect the results substantially. We verify that couples who retire sequentially did *not* have more volatile realizations of health and financial variables than those couples who retire simultaneously. This might have been expected if shocks displaced couples' plans to retire simultaneously.

Table 1 presents summary statistics for the variables we use. Note that we observe potential censoring months even for the observations that are uncensored in the data. This means that even though we assume that the censoring time is independent from retirement dates (conditional on the covariates), we do not need to model the distribution of censoring times to simulate the model. When drawing observations from the model to fit the auxiliary duration models, we are able to censor the simulated observations using the date when respondents were last interviewed or died as censoring dates even for those who retire earlier than that in the data (i.e., the "censoring month" in Table 1).

In Table 2, we present an overview of intra-household differences. Most of the couples marry within their own race but there is substantial variation in terms of education. Many couples report different health statuses, and accordingly, there is a substantial difference in health expenditures. There are also differences with respect to insurance and pension ownership. Figure 6 presents the Kaplan-Meier estimates for the retirement behavior in our sample (measured in months since the oldest partner turned 60 years old).

---

numbers. It is concave for positive values and convex for negative ones. In the computations, we also divide the transformed variable by 10^3 to avoid overflow.

Table 1: Summary statistics

Variable	All Observations Mean	N	Uncensored Mean	N	Censored Mean	N
Gender	0.50	2568	0.56	1265	0.44	1303
Min(Ret. Month, Cens. Month)	52.67	2568	44.43	1265	60.66	1303
Censored	0.51	2568	0.00	1265	1.00	1303
Censoring Month[a]	85.04	2568	110.15	1265	60.66	1303
Age Diff.	4.11	2454	3.80	1234	4.43	1220
Non-Hisp. White	0.79	2568	0.82	1265	5.07	1303
Non-Hisp. Black	0.11	2568	0.09	1265	0.12	1303
Other Race	0.02	2568	0.02	1265	0.03	1303
Hispanic	0.08	2568	0.06	1265	0.10	1303
< High School	0.16	2568	0.18	1265	0.15	1303
HS or GED	0.36	2568	0.39	1265	0.34	1303
Some College	0.24	2568	0.22	1265	0.26	1303
College or Above	0.23	2568	0.21	1265	0.25	1303
NE	0.18	2568	0.18	1265	0.04	1303
MW	0.26	2568	0.26	1265	0.25	1303
SO	0.40	2568	0.37	1265	0.43	1303
WE	0.16	2568	0.17	1265	0.15	1303
Health Insurance	0.88	2554	0.89	1257	0.86	1297
Very Good Health	0.56	2568	0.57	1265	0.85	1296
Good Health	0.30	2568	0.30	1265	0.30	1303
Poor Health	0.14	2568	0.13	1265	0.14	1303
Pension (DB)	0.26	2568	0.30	1265	0.22	1303
Pension (DC)	0.24	2568	0.21	1265	0.27	1303
Tot. Health Expen.[b]	8.22	2180	9.48	1220	6.61	960
Financial Wealth[c]	81.22	2568	88.76	1265	73.90	1303

a. For those uncensored, the censoring month is either the last interview or death date, whichever is the earlier date. It is used in the simulations for indirect inference.
b. Inflation-adjusted using the CPI to thousands of 2000 US dollars.

Table 2: Intra-Household Differences

	Prop. or Diff.	N of Couples
Same Race (proportion)	0.9502	1284
Same Education (proportion)	0.4618	1284
Same Self-Reported Health (proportion)	0.4938	1284
Health Insurance (both) (proportion)	0.8386	1270
Health Insurance (neither) (proportion)	0.0882	1270
DB Pension (both) (proportion)	0.0833	1284
DB Pension (neither) (proportion)	0.5600	1284
DC Pension (both) (proportion)	0.0685	1447
DC Pension (neither) (proportion)	0.5927	1284
Health Exp. (difference) (US$1,000)	3.0840	1090

Only couples with no missing variables. Inflation-adjusted health expenditures in Jan/ 2000 USD.

4 Results

We now present our estimation results using monthly data on couples' retirement. The discount rate ρ is set to 5% per year (i.e., 0.004 per month) and the threat points are set at 0.6 times the utility level an individual would have obtained without the retirement externality.[15] The number of simulations in each set of estimates is $R = 10$. Figure 6, which displays estimates for the marginal cumulative distribution functions of husbands and wives, suggests that a kink might be present, at least for men, around month twenty-four since the oldest household member turned 60. Since the oldest member is usually the husband, this corresponds to those turning 62 years old and becoming eligible for early retirement. In our model, we accommodate this time-varying variable by allowing Z to have a jump when the individual turns 62 years old. Since retirement is recorded in monthly intervals, we aggregate the retirement date from our simulations at the month level.

As is often the case in structural estimation, it can be difficult to understand what features of the data identify the parameters of the model. Within the duration literature, this has led to a sizeable number of papers dealing with semiparametric identification. This literature is useful in shedding light on the relation between the parameters of the model and the underlying data. First, note that the functions $Z_i(\cdot)$, $\varphi_i(\cdot)$ and the marginal distribution for K_i are formally identified (up to scale) if covariates have a support large enough so that $\varphi_j(x_j)$ can be made arbitrarily close to zero. For such an individual, it is essentially optimal to have $t_j = \infty$. The other spouse will then optimally retire at T_i such that

$$\log Z_i(T_i) = -\log \varphi_i(x_i) + \log K_i$$

and one can apply the arguments in Ridder (1990) to identify $Z_i(\cdot)$, $\varphi_i(\cdot)$ and the marginal distribution of K_i (up to scale). We note also that this identification argument operates irrespective of the values of A_1 and A_2 (or asymmetries in the bargaining power). Intuitively,

[15] In our estimations, we experimented with other multiples of this utility level as well. See the discussion at the end of this section.

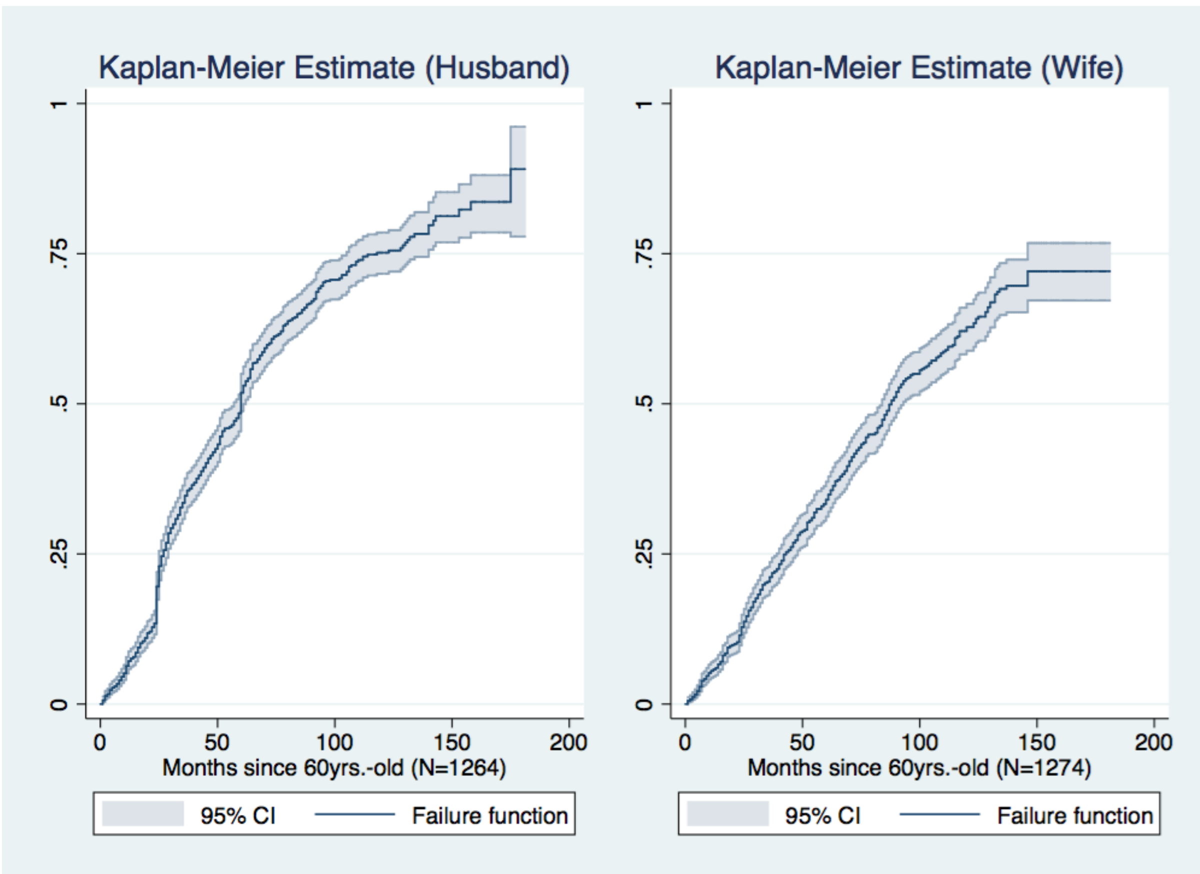

Figure 6: Kaplan-Meier Estimates: Husband and Wife

this argument would apply if the explanatory variables take values that make one of the spouses strongly attached to the labor force given his or her covariate values. In our data, for example, about 5% of the husbands who do not have a defined benefit pension plan retire after more than 126 months (10.5 years) since the oldest member of the household turned 60. Similarly, for the wives, 5% of those without a defined benefit pension plan retire more than 140 months (11.7 years) since the oldest member turned 60. (Our data are observed at the month level. The identification of (continuous-time) duration models under interval censoring is examined in detail, for instance, in Ridder (1990).)

Having identified $Z_i(\cdot), \varphi_i(\cdot)$ and the marginal distribution of K_i, the probability of joint retirement is driven by the interaction parameter δ. When $\delta = 1$ there are no retirement complementarities and joint retirement happens with zero probability. Larger values of δ will induce larger retirement complementarities, which should make joint retirement more likely. Even in the event of sequential retirement, whereas the first person to retire always retires at $Z_i^{-1}(K_i/\varphi_i)$, larger values of δ will lead to earlier retirement of the second person, providing additional variation to identify δ (see Appendix for details).

To understand why features of the joint distribution of K_1 and K_2 such as the τ in the copula are identified, consider a point (k_1, k_2). If the joint support of covariates is large enough, then for that point there is a pair (φ_1, φ_2) that induces sequential retirement in a neighborhood of (k_1, k_2). When there is sequential retirement, the retirement dates t_1 and t_2 are a one-to-one mapping from k_1 and k_2. For example, if $t_1 < t_2$, then t_1 is equal to $Z_1^{-1}(k_1/\varphi_1)$ and t_2 is also uniquely determined (see footnote 7). From the FOC, it is clear that, given (t_1, t_2) (and $k_1 = Z_1(t_1)\varphi_1$), one can uniquely retrieve the corresponding k_2. Since we have a one-to-one mapping, the joint distribution of (T_1, T_2) is therefore informative about the joint density of (K_1, K_2). A different distribution of (K_1, K_2) in the neighborhood of (k_1, k_2) changes the probability of (T_1, T_2) given the covariates corresponding to the initial choice of (φ_1, φ_2) leading to sequential retirement.

Tables 3 and 4 present our estimates. The results are very robust across covariate specifications. There is positive duration dependence: retirement is more likely as the household ages. Age differences tend to increase the retirement hazard for men and decrease it for women. Since men are typically older and we count “family age” from the 60th year of the older partner, a larger age difference implies that the wife is younger at time zero and less likely to retire at any “family age” than an older woman (i.e., a similar wife in a household with a lower age difference). Both non-white men and women have a lower retirement hazard than non-Hispanic whites. The hazard for a Hispanic man is between 0.53 ($= \exp(-0.63)$) and 0.63 ($= \exp(-0.46)$) of a white man's, for example.

Women with a high school diploma or GED tend to retire earlier than those without a high school degree, whereas women with some college or with a college education or above seem to retire later, but the coefficients on those categories are not statistically significant. For men, more educated husbands tend to retire later and the association is statistically significant for college-educated males. Husbands in the Northeast and West tend to retire earlier, whereas those in the South retire later than those in the Midwest but the coefficients are statistically insignificant in some specifications. Geographical region does not seem to play a statistically significant role for women either. Western and Southern wives do seem to retire earlier in all covariate specifications, but then again, standard errors are quite imprecise for most of the specifications.

Self-reported health lowers the hazard, with healthier people retiring later than those in poor health. Having health insurance increases the hazard for both husbands and wives, though not in a statistically significant way. Total health expenditures increase the hazard for female and for males (being statistically significant for the latter). Having a defined benefit pension plan increases the probability of retirement for both genders, but it is numerically and statistically much stronger for men. A defined contribution plan negatively affects the hazard for both, but here female effects are numerically and statistically more pronounced than those for men. Wealthier men and women tend to retire earlier and the effect is particularly more pronounced for women.

The interaction parameter ranges from 1.03 to 1.08 across our specifications. In terms of our model, this means that the utility flow of retirement increases by more than

Table 3: WIVES' Simultaneous Duration

Variable	**Coef.** (Std. Err.)	**Coef.** (Std. Err.)	**Coef.** (Std. Err.)	**Coef.** (Std. Err.)	**Coef.** (Std. Err.)	**Coef.** (Std. Err.)
δ	1.07	1.08	1.07	1.03	1.04	1.05
	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
θ_1	1.24	1.23	1.27	1.29	1.27	1.29
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Constant	-5.75 **	-5.71 **	-5.66 **	-5.93 **	-5.86 **	-6.08 **
	(0.21)	(0.27)	(0.31)	(0.35)	(0.35)	(0.36)
$\geq$ 62 yrs-old	8.40 †	11.27	12.65	9.98 †	8.24	8.17
	(4.31)	(8.85)	(9.64)	(5.94)	(6.57)	(5.93)
Age Diff.	-0.07 **	-0.07 **	-0.08 **	-0.08 **	-0.08 **	-0.08 **
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Non-Hisp. Black		-0.15	-0.16	-0.1	-0.11	-0.06
		(0.15)	(0.16)	(0.16)	(0.16)	(0.17)
Other race		-0.64 †	-0.63 †	-0.68 †	-0.68 †	-0.57 †
		(0.33)	(0.36)	(0.33)	(0.34)	(0.33)
Hispanic		-0.5 **	-0.54 **	-0.41 †	-0.41 †	-0.34
		(0.19)	(0.19)	(0.21)	(0.2)	(0.21)
High school or GED		0.03	0.08	0.13	0.14	0.13
		(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Some college		-0.13	-0.12	-0.13	-0.11	-0.17
		(0.17)	(0.17)	(0.17)	(0.18)	(0.18)
College or above		-0.06	-0.01	0.06	0.05	-0.08
		(0.19)	(0.19)	(0.2)	(0.2)	(0.2)
NE		-0.04	-0.01	-0.08	-0.1	-0.1
		(0.14)	(0.15)	(0.15)	(0.16)	(0.16)
SO		0.04	0.03	0.08	0.08	0.08
		(0.11)	(0.12)	(0.12)	(0.12)	(0.12)
WE		0.19	0.2	0.17	0.16	0.14
		(0.15)	(0.15)	(0.15)	(0.15)	(0.16)
V Good Health			-0.23	-0.15	-0.16	-0.16
			(0.15)	(0.17)	(0.17)	(0.17)
Good Health			-0.26	-0.2	-0.24	-0.23
			(0.16)	(0.18)	(0.17)	(0.18)
Health Insurance				0.11	0.15	0.18
				(0.15)	(0.15)	(0.15)
Tot. Health Exp.				0.13	0.13	0.13
				(0.09)	(0.08)	(0.08)
Pension (DC)					-0.23 †	-0.23 †
					(0.13)	(0.13)
Pension (DB)					0.02	0.05
					(0.11)	(0.11)
Fin. Wealth						0.75 **
						(0.21)
τ	0.42	0.34	0.46	0.62	0.51	0.45
	(0.36)	(0.35)	(0.38)	(0.29)	(0.26)	(0.26)
Function Value	0.28	0.44	0.55	1.15	1.18	1.07
Number of Obs.	1227	1227	1227	1037	1037	1037

Significance levels : † : 10% $*$: 5% $**$: 1%. Significance levels are not displayed for θ_1 or δ. Omitted categories are `Non-Hisp. White`, `Less than high school`, `Midwest or Other Region`, and `Poor Health` The threat point scale factor is 0.6, $\rho = 0.004$ and $R = 10$.

Table 4: HUSBANDS' Simultaneous Duration

Variable	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)	Coef. (Std. Err.)
δ	1.07	1.08	1.07	1.03	1.04	1.05
	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
θ_1	1.21	1.21	1.25	1.25	1.24	1.24
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	-5.75 **	-5.35 **	-5.43 **	-5.71 **	-5.64 **	-5.7 **
	(0.21)	(0.21)	(0.22)	(0.26)	(0.27)	(0.25)
$\geq$ 62 yrs-old	33.68 **	33.27 **	39.04 **	41.61 **	39.04 **	38.59 **
	(8.59)	(7.15)	(8.41)	(9.52)	(9.49)	(7.18)
Age Diff.	0.02 †	0.03 **	0.03 **	0.03 **	0.03 **	0.03 **
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Non-Hisp. Black		-0.18	-0.2	-0.25	-0.27 †	-0.26
		(0.16)	(0.15)	(0.16)	(0.16)	(0.16)
Other race		-0.18	-0.17	-0.09	-0.13	-0.16
		(0.28)	(0.28)	(0.3)	(0.28)	(0.29)
Hispanic		-0.63 **	-0.63 **	-0.52 **	-0.51 **	-0.46 †
		(0.18)	(0.18)	(0.18)	(0.18)	(0.18)
High school or GED		-0.09	-0.09	-0.05	-0.02	-0.02
		(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Some college		-0.33 †	-0.32 †	-0.3 †	-0.25 †	-0.25 †
		(0.14)	(0.13)	(0.14)	(0.14)	(0.14)
College or above		-0.5 **	-0.49 **	-0.47 **	-0.44 **	-0.46 **
		(0.12)	(0.13)	(0.13)	(0.13)	(0.14)
NE		0.04	0.07	0.09	0.11	0.12
		(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
SO		-0.24 †	-0.22 †	-0.16	-0.15	-0.15
		(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
WE		0.06	0.07	0.02	0.01	0.02
		(0.12)	(0.12)	(0.13)	(0.12)	(0.12)
V Good Health			-0.12	-0.08	-0.12	-0.11
			(0.13)	(0.14)	(0.14)	(0.14)
Good Health			-0.09	-0.04	-0.08	-0.08
			(0.14)	(0.14)	(0.15)	(0.14)
Health Insurance				0.18	0.11	0.12
				(0.13)	(0.13)	(0.13)
Tot. Health Exp.				0.11 †	0.10 †	0.12 †
				(0.05)	(0.05)	(0.05)
Pension (DC)					-0.1	-0.12
					(0.11)	(0.1)
Pension (DB)					0.29 **	0.28 **
					(0.1)	(0.1)
Fin. Wealth						0.24
						(0.19)
τ	0.42	0.34	0.46	0.62	0.51	0.45
	(0.36)	(0.35)	(0.38)	(0.29)	(0.26)	(0.26)
Function Value	0.28	0.44	0.55	1.15	1.18	1.07
Number of Obs.	1227	1227	1227	1037	1037	1037

Significance levels : † : 10% $*$: 5% $**$: 1%. Significance levels are not displayed for θ_1 or δ. Omitted categories are `Non-Hisp. White`, `Less than high school`, `Midwest or Other Region`, and `Poor Health` The threat point scale factor is 0.6, $\rho = 0.004$ and $R = 10$.

3% when one's partner retires. In terms of the effect on the hazard rate of retirement, this corresponds to between 9% and 24% of the effect of having a defined benefit plan for men. We also note that the copula parameter hovers above 0.5 in many of our specifications, yielding a Kendall's rank correlation coefficient of about 0.2. As explained previously, this correlation is potentially due to sorting or other commonalities.

To gauge the quantitative importance of the retirement externality, we also computed the marginal effect of assigning every man to a defined contribution pension plan compared to a defined benefit plan, holding everything else fixed.[16] This resulted in a 17.2-month change in the median uncensored retirement date for men. The simulated effect on the women was a change in the median uncensored retirement date of 0.57 month. In other words, the indirect effect on the women through the retirement externality is about 3.3% of the direct effect on the men. Given the large amount of censoring, one might argue that the median uncensored retirement date is not representative of the data we actually use. We therefore also compared the effect on the 25th percentile of uncensored retirement dates. Here, the direct effect on the husbands was 1.9 months, while the effect on the wives corresponded to about 15.4% of that.

We also added spousal variables as covariates to the last specification. Those variables were: dummies for "very good health" and "good health" and dummies for defined benefit and defined contribution pensions. For males, none of the coefficients on the spousal variables is statistically significant. For females, only the coefficient on a defined benefit pension plan for the spouse is statistically significant. The coefficient of the husband having a defined benefit plan on a woman's duration (0.40) is larger than that of the man himself having a defined benefit pension plan on his own duration to retirement, which is 0.27 once we include the spousal covariates. In contrast, the point estimate of the effect of a wife having a defined benefit pension plan on the man's duration is also positive but statistically insignificant. The absence of an effect of spousal health is in line with previous findings in the literature (e.g., Coile (2004a)). The effect of switching husbands from a defined contribution to a defined

[16]We used 100 simulation draws per individual to generate predictions under each of these two scenarios.

benefit plan is still in line with our previous results: median duration until retirement for men increases by 16.6 months, whereas median time to retirement for wives increases by 0.8 month, corresponding to 4.7% of the direct effect on husbands' duration. The corresponding numbers for the 25th percentile are an effect of 1.6 months for the males and 0.4 months for the females.

All of the estimation results presented here can be thought of as GMM results. Rather than working with the same moment conditions for all specifications, we always work with the moment conditions that come from the scores of the Weibull proportional hazard auxiliary model using the same explanatory variables as in the final model, combined with the scores from the pseudo-likelihood function for the ordered logit using only a constant as an explanatory variable. This means that the number of overidentifying restrictions is one for all of the specifications. As a specification test, we should therefore compare our minimized objective function to a Chi-squared distribution with one degree of freedom (see Proposition 2 and ensuing discussion in Smith (1993)). The p-values associated with this test of overidentifying restrictions range from 28% to 60%. The average p-value across the six specifications is 40%. This suggests that our specification provides a good fit to the moments implicitly used in the estimation and have good predictive power on the retirement behavior of couples. This is confirmed by comparing the Kaplan-Meier estimator of the observed durations to durations simulated using the sixth specification from Tables 3 and 4. This is reported in Figure 7. Figure 8 reports the same graphs after breaking the sample into two, depending on whether the health status is Very Good.

In Figure 9, we verify the robustness of our estimates to different threat point levels. As mentioned previously, we set $A_i, i = 1, 2$, equal to 0.6 of the utility spouse i would obtain without the utility externality from joint retirement. In the graph we plot 95% confidence intervals and point estimates of δ for various proportions of the utility one would get in case the partner were not to retire in the third specification from Tables 3 and 4.[17] As seen from

[17]The confidence intervals are asymmetric since lower and upper bounds correspond to symmetric lower and upper bounds on $\ln(\delta - 1)$ (which is the parameterization used in our computations for the estimation of δ).

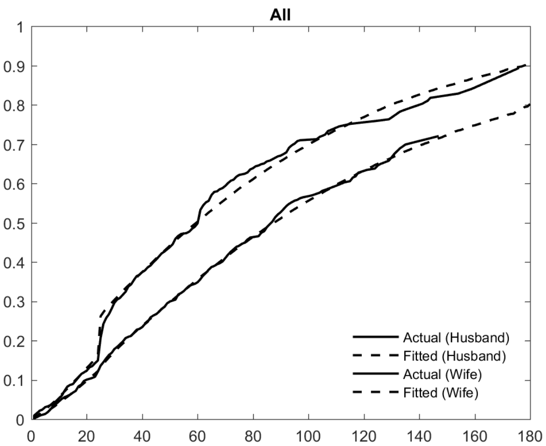

Figure 7: Predicted and Actual Distribution of Retirement Durations

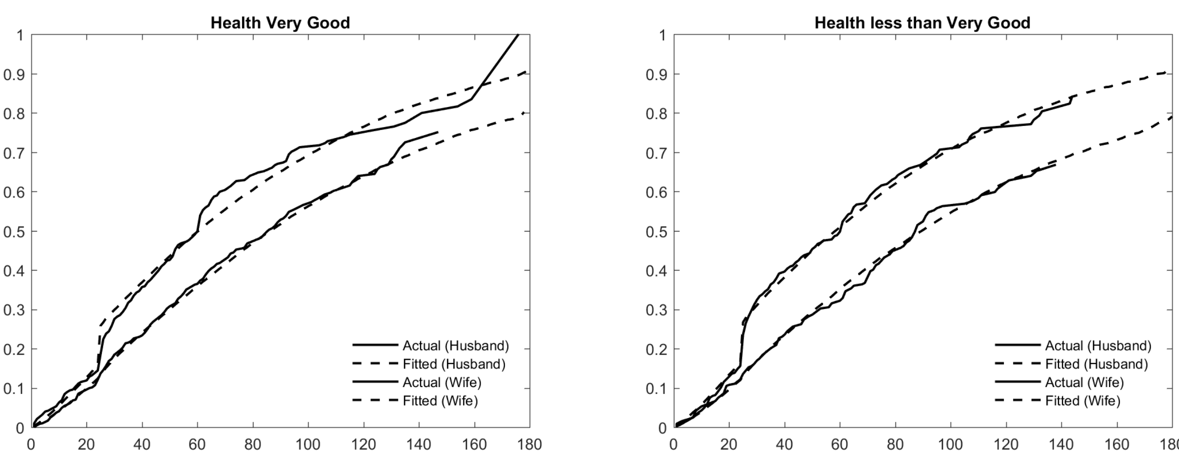

Figure 8: Predicted and Actual Distribution of Retirement Durations Conditional of Health

the figure, point estimates hover around an average of 1.067, which is essentially the estimate presented in our main tables (i.e., 1.070).

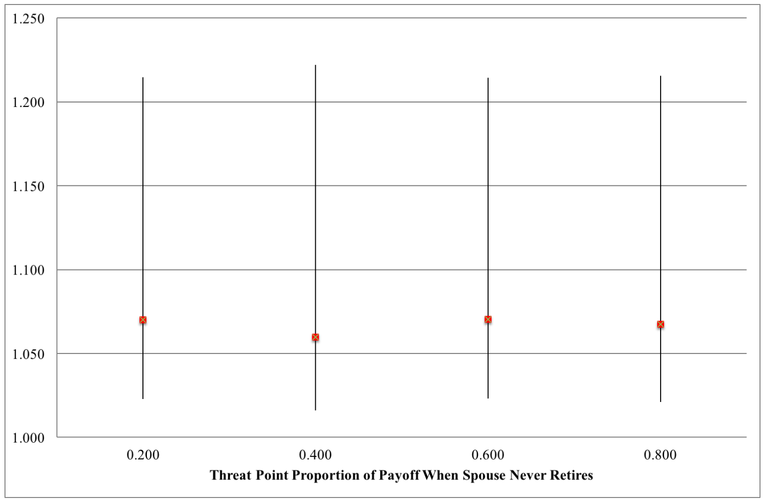

Figure 9: Robustness of $\hat{\delta}$ to Different Threat Point Specifications

Because the differential utility from joint retirement may depend on household characteristics, we also split our estimation into households where husband and wife are within 3 years apart in age and households where their age difference is greater than 3 years. The results are presented in Table 5 for the covariates used in the third specification from Tables 3 and 4. As expected, the interaction coefficient is higher for households closer in age. Interestingly, for households closer in age, Kendall's rank correlation coefficient ($\tau/(2+\tau)$) is lower (0.15 versus 0.18). Retirement timing for wives in both types of households responds more to health conditions than in the baseline specification in Table 3. More educated husbands tend to retire later in households farther apart in age when compared to our baseline results. The hazard is also comparatively higher for non-Hispanic black wives in couples farther apart in age than in the baseline specification. For households closer in age, Hispanic wives retire later than in the baseline specification and college seems to decrease the retirement hazard for wives while it raises it for households farther apart in age. The coefficient on `Other Race` is quite different across subsamples. The proportion of individuals of other race is nonetheless small in both: around 3.2% among couples more than 3 years apart in age and only 1.8% among those closer in age. For men, there is also heterogeneity for many of the coefficients. The indicator for early retirement eligibility carries a stronger coefficient for men in households closer in age, whereas it is statistically insignificant for wives in those households. On the other hand, the coefficient on early retirement eligibility for wives is significant and higher than in our baseline results when households are farther apart in age.

Finally, to evaluate whether joint retirement is likely to be an outcome from a common shock, as opposed to the interaction between husband and wife, we compare the time variation of our regressors across couples who retire simultaneously and couples who retire sequentially. For the survey waves preceding retirement of any member in the household, we look at the average proportional changes in financial assets and health expenditures and average changes in self-reported health status, pension plans (defined benefit and defined contribution) and health insurance. For all of these variables, couples retiring simultaneously displayed at least as much stability (if not more) in the survey waves preceding retirement as those retiring sequentially. For example, financial assets for those who end up retiring simultaneously are much more stable than for couples who retire sequentially: the average relative change in financial wealth across survey waves preceding retirement is a factor of 4.847 for those who retire simultaneously versus a factor of 10.850 for those who retire sequentially. Standard deviations were also lower for those couples retiring simultaneously. The same pattern arises even when the factor is deflated by the growth in the S&P500 stock market index. Furthermore, there is no discernible statistical difference between the average change in financial assets from survey wave to survey wave for these two groups. Consequently, it is unlikely that shocks to financial wealth (and, for that matter, that shocks to any of the variables listed above) explain the joint retirement decision in our sample.

5 Concluding Remarks

We have presented a new duration model that nests the usual generalized accelerated failure time model, but accounts for joint termination of a pair of spells in a way that is consistent with an economic model of joint decision making. The econometric model is based on a very

Table 5: Simultaneous Duration by Age Diff.

Variable	$\leq$ 3 yrs. Wife Coef. (Std. Err.)	$\leq$ 3 yrs. Husband Coef. (Std. Err.)	$>$ 3 yrs. Wife Coef. (Std. Err.)	$>$ 3 yrs. Husband Coef. (Std. Err.)
δ	1.11		1.07	
	(0.08)		(0.05)	
θ_1	1.27	1.23	1.29	1.22
	(0.09)	(0.07)	(0.07)	(0.06)
Constant	-5.34 **	-5.29 **	-5.74 **	-5.28 **
	(0.52)	(0.37)	(0.37)	(0.34)
$\geq$ 62 yrs-old	2.93	42.43 **	18.16 †	30.98 **
	(8.74)	(15.58)	(9.63)	(9.63)
Age Diff.	-0.14 **	0.01	-0.01	0.11 **
	(0.03)	(0.01)	(0.02)	(0.02)
Non-Hisp. Black	-0.51 †	-0.4 †	0.01	0.13
	(0.28)	(0.19)	(0.19)	(0.21)
Other race	0.32	-0.08	-1.13 **	-0.33
	(0.42)	(0.3)	(0.43)	(0.58)
Hispanic	-0.63 †	-0.49 †	-0.43 †	-0.7 †
	(0.33)	(0.23)	(0.26)	(0.28)
High school or GED	-0.13	-0.11	0.3	-0.05
	(0.27)	(0.16)	(0.2)	(0.19)
Some college	-0.14	-0.31 †	-0.05	-0.13
	(0.28)	(0.17)	(0.22)	(0.21)
College or above	-0.32	-0.5 **	0.33	-0.37 †
	(0.33)	(0.18)	(0.22)	(0.2)
NE	-0.07	-0.05	0.04	0.2
	(0.26)	(0.18)	(0.17)	(0.16)
SO	0.38 †	-0.09	-0.02	-0.34 †
	(0.2)	(0.15)	(0.15)	(0.15)
WE	0.38	0.04	0.18	0.11
	(0.28)	(0.17)	(0.17)	(0.17)
V Good Health	-0.31	-0.11	-0.31	-0.16
	(0.23)	(0.18)	(0.2)	(0.21)
Good Health	-0.49 †	0.04	-0.41 †	-0.3
	(0.25)	(0.18)	(0.21)	(0.22)
τ	0.34		0.43	
	(0.47)		(0.47)	
Function Value	0.16		0.07	
Number of Obs.	590		637	

Significance levels :† : 10% $*$: 5% $**$: 1%. Significance levels are not displayed for θ_1 or δ. Omitted categories are `Non-Hisp. White`, `Less than high school`, `Midwest or Other Region`, and `Poor Health`. The threat point scale factor is 0.6, $\rho = 0.004$ and $R = 5$.

simple economic model with Nash bargaining and it can generate concurrent termination of spells with positive probability as well as interdependence between the durations when they are not concurrent, even when the underlying unobservables are independent.

We then applied the model to the retirement of husband and wife using data from the Health and Retirement Study. The main empirical finding is that simultaneity seems economically important. Since the econometric model is based on a simple economic model, it is possible to interpret the estimates in terms of the underlying preferences. In our preferred specification, the indirect utility associated with being retired increases by approximately 5% if one's spouse is already retired. By comparison, a defined benefit pension plan increases indirect utility by 32%. The estimated model also predicts that the marginal effect of a change in the husbands' pension plan on wives' retirement dates is about 3-25% of the direct effect on the husbands'.

References

An, M., B. Christensen, and N. Gupta (2004): "Multivariate Mixed Proportional Hazard Modeling of the Joint Retirement of Married Couples," *Journal of Applied Econometrics*, 19(6), 687–704.

Banks, J., R. Blundell, and M. Casanova Rivas (2007): "The Dynamics of Retirement Behavior in Couples: Reduced-Form Evidence from England and the U.S.," Working Paper, University College London (UCL).

Binmore, K., A. Rubinstein, and A. Wolinsky (1986): "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics*, 17, 176–188.

Blau, D. (1998): "Labor Force Dynamics of Older Married Couples," *Journal of Labor Economics*, 16(3), 595–629.

Blau, D., and D. Gilleskie (2004): "Health Insurance and Retirement of Married Couples," Working Paper, University of North Carolina at Chapel Hill.

Browning, M., P.-A. Chiappori, and V. Lechene (2006): "Collective and Unitary Models: A Clarification," *Review of Economics of the Household*, 4(1), 5–14.

Burtless, G. (1990): "Comment on M. Hurd's 'The Joint Retirement Decisions of Husbands and Wives'," in *Issues in the Economics of Aging*, ed. by D. A. Wise. Chicago: University of Chicago Press.

Casanova Rivas, M. (2010): "Happy Together: A Structural Model of Couples' Joint Retirement Choices," Working Paper, University of California at Los Angeles (UCLA).

Chiappori, P.-A. (1992): "Collective Labor Supply and Welfare," *Journal of Political Economy*.

Chiappori, P.-A., O. Donni, and I. Komunjer (2012): "Learning from a Piece of Pie," *Review of Economic Studies*, 79(1), 162–195.

Clayton, D., and J. Cuzick (1985): "Multivariate Generalizations of the Proportional Hazards Model," *Journal of the Royal Statistical Society Series B*, 34, 187–220.

Coile, C. (2004a): "Health Shocks and Couples' Labor Supply Decisions," NBER Working Paper 10810.

——— (2004b): "Retirement Incentives and Couples' Retirement Decisions," *Topics in Economic Analysis and Policy*, 4(1).

de Paula, A. (2009): "Inference in a Synchronization Game with Social Interactions," *Journal of Econometrics*, 148(1), 56–71.

——— (2016): "Econometrics of Network Models," CeMMAP Working Paper, CWP06/16.

Gallant, R., and G. Tauchen (1996): "Which Moments to Match?," *Econometric Theory*, 12(4).

Gourieroux, C., and A. Monfort (1996): *Simulation-Based Econometric Methods*. Oxford: Oxford University Press.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Gustman, A., and T. Steinmeier (2000): "Retirement in Dual-Career Families: A Structural Model," *Journal of Labor Economics*, 18(3), 503–545.

——— (2004): "Social Security, Pensions and Retirement Behaviour with the Family," *Journal of Applied Econometrics*, 19, 723–737.

——— (2009): "Integrating Retirement Models," NBER Working Paper 15607.

Hahn, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61(4), 607–629.

Honoré, B., and A. de Paula (2010): "Interdependent Durations," *Review of Economic Studies*, 77, 11381163.

Hurd, M. (1990): "The Joint Retirement Decisions of Husbands and Wives," in *Issues in the Economics of Aging*, ed. by D. A. Wise. Chicago: University of Chicago Press.

Lancaster, T. (1990): *Econometric Analysis of Transition Data*. Cambridge University Press.

Magnac, T., J.-M. Robin, and M. Visser (1995): "Analysing Incomplete Individual Employment Histories Using Indirect Inference," *Journal of Applied Econometrics*, 10, S153–S169.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–542.

Marshall, A., and I. Olkin (1967): "A Multivariate Exponential Model," *Journal of the American Statistical Association*, 62, 30–44.

Michaud, P.-C. (2003): "Labor Force Participation Dynamics and Social Security Claiming Decisions of Elderly Couples," Working Paper, CentER Tilburn University and IZA.

Michaud, P.-C., and F. Vermeulen (2011): "A Collective Labor Supply Model with Complementarities in Leisure: Identification and Estimation by Means of Panel Data," *Labour Economics*, 18, 159–167.

Nash, J. (1950): "The Bargaining Problem," *Econometrica*, 18, 155–162.

Ridder, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57.

Smith, A. (1993): "Estimating Nonlinear Time Series Models Using Vector-Autoregressions: Two Approaches," *Journal of Applied Econometrics*, 8, S63–S84.

Trivedi, P. K., and D. M. Zimmer (2006): "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, 1(1), 1–111.

Zeuthen, F. (1930): *Problems of Monopoly and Economic Warfare*. London: Routledge.

Appendix

Computational Details

The sample moment conditions implied by the auxiliary model used for indirect inference in this paper are discontinuous functions of the structural parameters. We calculate the minimizer of the corresponding GMM minimization problem as follows.

1. δ is parameterized as $\exp\left(\widetilde{\delta}\right)+1$; θ_1, θ_2, τ, and the jumps in $Z(\cdot)$ are parameterized as $\exp\left(\widetilde{\theta}_1\right)$, $\exp\left(\widetilde{\theta}_2\right)$, $\exp\left(\widetilde{\tau}\right)$, $\exp\left(\alpha_1\right)$ and $\exp\left(\alpha_2\right)$.

2. Weibull models are estimated separately for husbands and wives as part of the auxiliary model. The estimates from this are the starting values for the θ's and β's. The starting values for δ and τ are 1.08 and $\exp(-1)$. The starting values for the jumps are $\exp(1)$ and $\exp(3.5)$ for females and males, respectively. The starting values for the objective functions for Specifications 1-6 range from 50.6 to 60.2.

3. The parameters are estimated by particle swarm using the built-in Matlab routine. The objective functions for Specifications 1-6 ranged from 0.27 to 1.64 after this.

4. The following loop of procedures was used until a loop produced a change in the parameter estimate of less than 10^{-5}. (The number of loops was restricted to be between 5 and 20.)

   (a) particle swarm using the built-in Matlab routine

   (b) Powell's conjugate direction method

   (c) downhill simplex using Matlab's `fminsearch` routine

   (d) pattern search using Matlab's built-in routine

   (e) particle swarm focusing on the jump-parameters using the build-in Matlab routine

5. Estimation of the asymptotic variance of the indirect inference estimator requires estimation of the variance of the element in the moment condition as well as estimation of

the derivative of its expectation. The latter is calculated by a numeric derivative after increasing the number of simulation replications by a factor of 20. For the step-size in the numeric derivative, we choose 0.01, 0.02, ..., 0.09, 0.1, and report the median of the implied estimated standard errors. Table 6 reports the reported standard errors for Specification 3 along with the standard errors associated with the different step-sizes. Table 6 suggests that the reported standard errors are not too sensitive to the way we choose the bandwidth. This is an important advantage of increasing the number of simulation draws in the estimation of the standard errors.

The discount factor is fixed and not estimated.

Table 6: **The Effect of Bandwidth on the Reported Standard Errors**

Females

Bandwidth	Median	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
δ	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
θ_1	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
$\geq$ 62 yrs-old	6.82	5.00	5.55	6.49	6.28	6.78	6.93	7.12	7.22	6.87	7.11
Constant	0.30	0.30	0.28	0.28	0.28	0.29	0.30	0.31	0.31	0.31	0.32
Age Diff.	0.16	0.16	0.16	0.16	0.16	0.16	0.15	0.15	0.16	0.16	0.16
Non-Hisp. Black	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.16	0.15
Other race	0.35	0.31	0.35	0.36	0.37	0.37	0.36	0.34	0.34	0.34	0.34
Hispanic	0.19	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.20	0.20
High school or GED	0.16	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Some college	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
College or above	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
NE	0.15	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
SO	0.11	0.11	0.11	0.12	0.12	0.11	0.11	0.11	0.11	0.11	0.11
WE	0.15	0.14	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
V Good Health	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Good Health	0.16	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
τ	0.35	0.31	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.34	0.35

Males

Bandwidth	Median	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
δ	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
θ_1	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04
$\geq$ 62 yrs-old	7.86	5.79	5.99	6.72	7.06	7.80	8.47	8.17	8.52	8.35	7.93
Constant	0.23	0.23	0.24	0.24	0.23	0.23	0.22	0.22	0.23	0.23	0.23
Age Diff.	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.07	0.07
Non-Hisp. Black	0.15	0.16	0.16	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.15
Other race	0.28	0.29	0.30	0.28	0.28	0.28	0.28	0.28	0.28	0.27	0.28
Hispanic	0.18	0.18	0.18	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18
High school or GED	0.12	0.12	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Some college	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
College or above	0.13	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.13
NE	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
SO	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
WE	0.12	0.13	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
V Good Health	0.13	0.14	0.14	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Good Health	0.14	0.14	0.14	0.14	0.14	0.14	0.13	0.14	0.14	0.14	0.14
τ	0.35	0.31	0.35	0.35	0.35	0.35	0.35	0.34	0.34	0.34	0.35

This table presents the estimated standard errors using different step-sizes to calculate the numeric derivative. The first column (Median) presents the median of the estimated standard error for each parameter.

Appendix for Referees

AUXILIARY MODELS FOR INDIRECT INFERENCE

Log-likelihood Derivatives: Weibull Model

$$\frac{\partial \log \mathcal{L}_i}{\partial \alpha_i} = \sum_{n=1}^{N}(1-c_{i,n})\left(\frac{1}{\alpha_i}+\log\left(t_{i,n}\right)\right) - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)\exp\left(x_{i,n}'\beta_i\right)$$

$$\frac{\partial \log \mathcal{L}_i}{\partial \beta_i} = \sum_{n=1}^{N}(1-c_{i,n})x_{i,n} - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\exp\left(x_{i,n}'\beta_i\right)x_{i,n}$$

$$\frac{\partial^2 \log \mathcal{L}_i}{\partial \alpha_i^2} = -\sum_{n=1}^{N}(1-c_{i,n})\frac{1}{\alpha_i^2} - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)^2\exp\left(x_{i,n}'\beta_i\right)$$

$$\frac{\partial^2 \log \mathcal{L}_i}{\partial \alpha_i \partial \beta_i'} = -\sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)\exp\left(x_{i,n}'\beta_i\right)x_{i,n}$$

$$\frac{\partial^2 \log \mathcal{L}_i}{\partial \beta_i \partial \beta_i'} = -\sum_{n=1}^{N} t_{i,n}^{\alpha_i}\exp\left(x_{i,n}'\beta_i\right)x_{i,n}x_{i,n}'$$

To impose $\alpha_i > 0$ in our computations we parameterize $\alpha_i = \exp\left(\theta\right)$. Then,

$$\frac{\partial \log \mathcal{L}_i}{\partial \theta} = \frac{\partial \log \mathcal{L}_i}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \theta} = \left(\sum_{n=1}^{N}(1-c_{i,n})\left(\frac{1}{\alpha_i}+\log\left(t_{i,n}\right)\right) - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)\exp\left(x_{i,n}'\beta_i\right)\right)\alpha_i$$

$$\frac{\partial \log \mathcal{L}_i}{\partial \beta_i} = \sum_{n=1}^{N}(1-c_{i,n})x_{i,n} - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\exp\left(x_{i,n}'\beta_i\right)x_{i,n}$$

$$
\begin{aligned}
\frac{\partial^2 \log \mathcal{L}_i}{\partial \theta^2} &= \frac{\partial}{\partial \theta}\left(\frac{\partial \log \mathcal{L}_i}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \theta}\right) \\
&= \frac{\partial^2 \log \mathcal{L}_i}{\partial \alpha_i^2}\left(\frac{\partial \alpha_i}{\partial \theta}\right)^2 + \frac{\partial \log \mathcal{L}}{\partial \alpha_i}\frac{\partial^2 \alpha_i}{\partial \theta^2} \\
&= \left(-\sum_{n=1}^{N}(1-c_{i,n})\frac{1}{\alpha_i^2} - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)^2 \exp\left(x_{i,n}'\beta_i\right)\right)\alpha_i^2 \\
&\quad - \left(\sum_{n=1}^{N}(1-c_{i,n})\left(\frac{1}{\alpha_i}+\log\left(t_{i,n}\right)\right) - \sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)\exp\left(x_{i,n}'\beta_i\right)\right)\alpha_i \\
\frac{\partial^2 \log \mathcal{L}_i}{\partial \theta \partial \beta_i'} &= \frac{\partial^2 \log \mathcal{L}_i}{\partial \alpha_i \partial \beta_i'}\frac{\partial \alpha_i}{\partial \theta} = \left(-\sum_{n=1}^{N} t_{i,n}^{\alpha_i}\log\left(t_{i,n}\right)\exp\left(x_{i,n}'\beta_i\right)x_{i,n}\right)\alpha_i \\
\frac{\partial^2 \log \mathcal{L}_i}{\partial \beta_i \partial \beta_i'} &= -\sum_{n=1}^{N} t_{i,n}^{\alpha_i}\exp\left(x_{i,n}'\beta_i\right)x_{i,n}x_{i,n}'
\end{aligned}
$$

Pseudo-likelihood Derivatives: Ordered Model

$$
\frac{\partial \mathcal{Q}}{\partial \gamma} = \sum_n \left[\left(1\left\{y_n \neq 0\right\} - \Lambda\left(x_{0n}'\gamma\right)\right)x_{0n} + \left(1\left\{y_n = 2\right\} - \Lambda\left(x_{1n}'\gamma\right)\right)x_{1n}\right]
$$

$$
\frac{\partial^2 \mathcal{Q}}{\partial \gamma \partial \gamma^\top} = -\sum_n \left[\left(\left(1-\Lambda\left(x_{0n}'\gamma\right)\right)\Lambda\left(x_{0n}'\gamma\right)\right)x_{0n}x_{0n}' + \left(\left(1-\Lambda\left(x_{1n}'\gamma\right)\right)\Lambda\left(x_{1n}'\gamma\right)\right)x_{1n}x_{1n}'\right]
$$

ADDITIONAL DERIVATIONS FOR IDENTIFICATION DISCUSSION

Here we provide details for the effect of δ on the retirement date of the second spouse to retire when there is sequential retirement. First, note that when $t_1 \approx 0$, applying the Implicit Function Theorem to the FOC for t_2 (see equation (2)) gives

$$
\frac{dt_2}{d\delta} = -\left[\frac{\frac{\partial^2 I}{\partial t_2 \partial \delta}\times(II) + \frac{\partial II}{\partial \delta}\times\frac{\partial I}{\partial t_2} + \frac{\partial^2 II}{\partial t_2 \partial \delta}\times(I) + \frac{\partial I}{\partial \delta}\times\frac{\partial II}{\partial t_2}}{\frac{\partial^2 I}{\partial^2 t_2}\times(II) + \frac{\partial II}{\partial t_2}\times\frac{\partial I}{\partial t_2} + \frac{\partial^2 II}{\partial^2 t_2}\times(I) + \frac{\partial I}{\partial t_2}\times\frac{\partial II}{\partial t_2}}\right], \qquad (7)
$$

where (I) and (II) are defined as in equation (2). The various terms can be signed as shown below:

$$
\begin{aligned}
&\tfrac{\partial I}{\partial \delta} = \varphi_1 \tilde{Z}_1(t_2) > 0 && \tfrac{\partial II}{\partial \delta} = \varphi_2 \tilde{Z}_2(t_2) > 0\\
&\tfrac{\partial I}{\partial t_2} = Z_1(t_2)e^{-\rho t_2}\varphi_1(1-\delta) < 0 && \tfrac{\partial II}{\partial t_2} = k_2 e^{-\rho t_2} - Z_2(t_2)\varphi_2\delta e^{-\rho t_2} > 0\\
&\tfrac{\partial^2 I}{\partial t_2 \partial \delta} = -Z_1(t_2)e^{-\rho t_2}\varphi_1 < 0 && \tfrac{\partial^2 II}{\partial t_2 \partial \delta} = k_2 e^{-\rho t_2} - Z_2(t_2)\varphi_2\delta e^{-\rho t_2} > 0\\
&\tfrac{\partial^2 I}{\partial t_2^2} = Z_1'(t_2)e^{-\rho t_2}\varphi_1(1-\delta) < 0 && \tfrac{\partial^2 II}{\partial t_2^2} = -\rho e^{-\rho t_2}(k_2 - Z_2(t_2)\varphi_2\delta) - Z_2'(t_2)e^{-\rho t_2} < 0.
\end{aligned}
$$

These and the fact that $(I) \geq 0$ and $(II) \geq 0$ imply that the denominator in expression (7) is *strictly* negative. To see that the numerator is also negative notice that

$$\lim_{\delta \to 1} \left[\frac{\partial II}{\partial \delta} \times \frac{\partial I}{\partial t_2} + \frac{\partial I}{\partial \delta} \times \frac{\partial II}{\partial t_2} \right] = \varphi_1 \tilde{Z}_1(t_2)[k_2 - Z_2(t_2)\varphi_2] = 0,$$

where the last equality follows because $k_2 = Z_2(t_2)\varphi_2$ at the optimally chosen t_2 when $\delta = 1$. Since

$$\frac{\partial}{\partial \delta} \left[\frac{\partial II}{\partial \delta} \times \frac{\partial I}{\partial t_2} + \frac{\partial I}{\partial \delta} \times \frac{\partial II}{\partial t_2} \right] = -\varphi_1\varphi_2 \left(Z_1(t_2)\tilde{Z}_2(t_2) + Z_2(t_2)\tilde{Z}_1(t_2) \right) e^{-\rho t_2} < 0,$$

it follows that

$$\frac{\partial II}{\partial \delta} \times \frac{\partial I}{\partial t_2} + \frac{\partial I}{\partial \delta} \times \frac{\partial II}{\partial t_2} < 0.$$

The other two remaining terms in the numerator are negative, which then implies that the numerator is negative. Consequently, (7) is negative: larger values of δ lead to earlier retirement by the second agent (i.e., lower t_2). Because $t_1 \approx 0$, I (and, consequently, t_2) will not depend on k_1. Having identified $Z_i(\cdot)$, $\varphi_i(\cdot)$ and the marginal distribution of K_2, this allows one to identify δ.