

CAF - WORKING PAPER #2025/16

# Estimating Choice Models with Piecewise Smooth Objective Functions: Application to Joint Retirement

Tatiana Rosá<sup>1</sup> | Siqi Wei<sup>2</sup>

<sup>1</sup>CAF Development Bank.  
[trosa@caf.com](mailto:trosa@caf.com)

<sup>2</sup>IE University: [siqi.wei@ie.edu](mailto:siqi.wei@ie.edu)

We study choice models with piecewise smooth objective functions and provide conditions under which introducing latent variables derived from regional components yields a censored-model-like representation. These latent variables can then be treated as potential outcomes, enabling tractable estimation via the stochastic EM algorithm. We illustrate the framework using two examples: responses to taxes and the  $(S, s)$  model. We further estimate a joint retirement decision model for European couples using an interdependent duration framework with our methods. The results provide empirical evidence of complementarities between spouses in the retirement process in Europe.

## KEYWORDS

Choice models, Piecewise smooth, Censored models, Latent variables, Joint retirement

---

Small sections of text that are less than two paragraphs may be quoted without explicit permission as long as this document is acknowledged. Findings, interpretations and conclusions expressed in this publication are the sole responsibility of its author(s) and cannot be, in any way, attributed to CAF, its Executive Directors or the countries they represent. CAF does not guarantee the accuracy of the data included in this publication and is not, in any way, responsible for any consequences resulting from its use.

**CAF - DOCUMENTO DE TRABAJO #2025/16**

Esta versión: 16 de diciembre 2025

# Estimación de modelos de elección con funciones objetivo suavemente definidas por tramos: aplicación a la jubilación conjunta

Tatiana Rosá<sup>1</sup> | Siqui Wei<sup>2</sup>

<sup>1</sup>CAF Development Bank.  
[trosa@caf.com](mailto:trosa@caf.com)

<sup>2</sup>IE University: [siqui.wei@ie.edu](mailto:siqui.wei@ie.edu)

Estudiamos modelos de elección con funciones objetivo suavemente definidas por tramos y presentamos condiciones bajo las cuales la introducción de variables latentes derivadas de componentes regionales produce una representación análoga a la de un modelo censurado. Estas variables latentes pueden tratarse como resultados potenciales, lo que permite una estimación tratable mediante el algoritmo EM estocástico. Ilustramos el marco con dos ejemplos: respuestas a impuestos y el modelo  $(S, s)$ . Además, estimamos un modelo de decisión conjunta de jubilación para parejas europeas utilizando un marco de duraciones interdependientes con nuestros métodos. Los resultados aportan evidencia empírica de complementariedades entre cónyuges en el proceso de jubilación en Europa.

#### KEYWORDS

Modelos de elección, Suavidad por tramos, Modelos censurados, Variables latentes, Jubilación conjunta

---

Pequeñas secciones del texto, menores a dos párrafos, pueden ser citadas sin autorización explícita siempre que se cite el presente documento. Los resultados, interpretaciones y conclusiones expresados en esta publicación son de exclusiva responsabilidad de su(s) autor(es), y de ninguna manera pueden ser atribuidos a CAF, a los miembros de su Directorio Ejecutivo o a los países que ellos representan. CAF no garantiza la exactitud de los datos incluidos en esta publicación y no se hace responsable en ningún aspecto de las consecuencias que resulten de su utilización.

## 1 | INTRODUCTION

Many structural models in applied economics are defined by optimization problems whose objective functions are piecewise smooth (Pudney, 1989). Such piecewise smooth features naturally arise from nonlinear budget constraints (such as kinked budget constraints) or nonsmooth operators embedded in the objective (such as max and min operators or fixed adjustment costs). These piecewise features are central in a broad set of applications—including taxation and bunching, durable (S, s) adjustment, and household or firm problems with nonconvexities—yet they complicate both solution and estimation.

The difficulty is not only numerical: nondifferentiability makes the model challenging to solve. When the optimum occurs at a corner or at a boundary where multiple regions meet, the mapping from observed choices to unobservables can become one-to-many: the same observed choice may be consistent with a range of latent shocks. In that sense, these problems resemble censored or incomplete-data models, even when the original model is written as a standard choice problem. This feature complicates likelihood-based estimation, since the likelihood contribution of boundary observations typically requires integrating over the set of latent shocks compatible with the observed choice. A common practice is to solve the model region by region—obtaining local optima in each piece and selecting the global optimum—paired with simulation-based estimation methods such as indirect inference (Smith Jr, 1993; Gouriéroux and Monfort, 1997). However, this approach requires specifying an auxiliary model, and identification tends to be less transparent, with statistical and computational efficiency depending on the choice of auxiliary specification and optimization method.

This paper develops an alternative route. We provide conditions under which a broad class of optimization-based choice models with piecewise smooth objective functions can be re-expressed in a reduced form that resembles an artificial censored model. This censored representation brings the problem into the domain of incomplete-data models, allowing us to apply estimation tools such as the stochastic EM (SEM) algorithm.

The key step is to introduce a finite set of well-defined latent variables that link the optimal choice to observables and unobservables. Concretely, these latent variables are constructed as (i) unconstrained global maximizers of the auxiliary functions (piecewise regional components) over the *entire* domain and (ii) maximizers at the boundaries between regions. Under our conditions—most importantly, global unimodality of each smooth component over an appropriate open extension of its region—any optimal choice for the original piecewise objective must coincide with one element of this finite set. Moreover, the observed choice data allow the researcher to infer which element of the latent set is selected. Consequently, these introduced latent variables can be interpreted as "practical" potential options offered to each unit and they select the optimal one from the finite list. This transformation effectively reframes the original optimization problem into a potential outcome framework.

On the one hand, from an optimization perspective, the transformation decomposes the global piecewise maximization into a finite comparison across a set of well-defined global and boundary solutions—rather than regional subproblems—providing an alternative representation of the optimization problem.

More importantly, the potential outcome framework facilitates alternative estimation methods. We propose using the stochastic EM (SEM) algorithm (Dempster et al., 1977; Celeux et al., 1996), an established method for models with incomplete or censored data. Leveraging the potential outcome framework, SEM alternates between an E-step, where we draw values for the introduced latent variables from the posterior distribution given the observables, and an M-step, where we estimate the potential outcome framework with

the E-step draws, until the parameters converge to a stationary distribution. The algorithm replaces integration with tractable iterative updates and has the potential to achieve fast convergence when combined with acceleration techniques (Wei, 2024). Two illustrative examples are discussed: 1) responses to taxes and transfers under progressive income taxation (Saez, 2010), and 2) the  $(S, s)$  model for infrequent adjustment (Attanasio, 2000; Bover, 2010).

This paper characterizes a class of optimization-based choice models with piecewise smooth objective functions that can be transformed into a reduced form resembling censored models. We develop an alternative solution method and propose the stochastic Expectation–Maximization (SEM) algorithm for their estimation.

Piecewise smooth objective functions in choice models can arise from various factors, such as bounded choice sets (e.g., non-negativity restrictions), nonlinear constraints (e.g., progressive tax systems), or objective functions containing nonlinear expressions of the choice variable (e.g., max and min operators) (Pudney, 1989). Estimating such models, which often requires solving them, can be challenging due to the frequent occurrence of nonlinearity and non-differentiability, which not only complicate the solution method but also can lead to data censorship. A potential approach involves sequentially obtaining local optima in each region and then selecting the global optimum, paired with simulation-based estimation methods such as indirect inference (Smith Jr, 1993; Gouriéroux and Monfort, 1997). However, this approach requires specifying an auxiliary model, and identification tends to be less transparent, with statistical and computational efficiency depending on the choice of auxiliary specification and optimization method.

In this paper, we provide conditions under which choice models with piecewise smooth objective functions can be transformed into a reduced form resembling an artificial censored model: while some choices are observed directly, there are censored outcomes consistent with a range of error terms, creating a one-to-many relationship between observables and unobservables. This transformation not only provides an alternative solution method to the choice model but also facilitates the application of the SEM algorithm, which is particularly useful for handling incomplete data.

The key to the transformation is the introduction of a set of well-defined latent variables that link the optimal choice to the relevant observables and unobservables. Specifically, the latent variables that we propose to introduce are the unconstrained global maximizers of the auxiliary functions (piecewise regional components) over the *entire* domain, as well as their maximizers at the boundaries between regions. The definition of these latent variables depends on the auxiliary functions (e.g., through first-order conditions) and thus depends on other observables, unobservables, and parameters of interest.

We show that when the auxiliary functions have a unique global maximizer with no local maxima over the entire domain, the optimal choice that maximizes the piecewise smooth objective function must correspond to one of the introduced latent variables. Moreover, researchers can infer which latent variable corresponds to the optimal choice based on the observed choice data. Consequently, these introduced latent variables can be interpreted as "practical" potential options offered to each unit and they select the optimal one from the finite list. This transformation effectively reframes the original optimization problem into a potential outcome framework.

On the one hand, from an optimization perspective, the transformation decomposes the global piecewise maximization into a finite comparison across a set of well-defined global and boundary solutions—rather than regional subproblems—providing an alternative representation of the optimization problem.

More importantly, the potential outcome framework facilitates alternative estimation methods. While the transformation applies generally, it is particularly useful in cases where

some observed choices are consistent with a range of unobservables—typically arising at corner or boundary solutions—creating a one-to-many mapping between choices and unobservables and thereby resembling a censored model. In such settings, direct estimation methods such as maximum likelihood require integrating over the entire feasible range of unobservables for censored observations, which is often computationally demanding or infeasible.

In such cases, we propose using the stochastic Expectation–Maximization (SEM) algorithm (Dempster et al., 1977; Celeux et al., 1996), an established method for models with incomplete or censored data. Leveraging the potential outcome framework, SEM alternates between an E-step, where we draw values for the introduced latent variables from the posterior distribution given the observables, and an M-step, where we estimate the potential outcome framework with the E-step draws, until the parameters converge to a stationary distribution. The algorithm replaces integration with tractable iterative updates and has the potential to achieve fast convergence when combined with acceleration techniques (Wei, 2024). Two illustrative examples are discussed: 1) responses to taxes and transfers under progressive income taxation (Saez, 2010), and 2) the  $(S, s)$  model for infrequent adjustment (Attanasio, 2000; Bover, 2010).

Finally, as an empirical application, we estimate the interdependent duration model of Honoré and de Paula (2018) with our method to study the joint retirement decisions among European couples. In this model, the retirement timing decision is determined through within-household Nash bargaining, considering potential complementarities in retirement decisions. The objective function of the bargaining problem has the piecewise feature, as the complementarity depends on the timing of the last person retiring, introducing a max operator into the objective function. This results in data "censoring" for jointly retiring couples, complicating both solving and estimating the model.

Applying the method, we introduce three pairs of well-defined latent variables as potential outcomes. The final choice of each household, observed by researchers, must correspond to one of these three options. This effectively transforms the original Nash bargaining problem into a potential outcome framework, which we leverage to develop the SEM estimation method. The proposed algorithm demonstrates numerical stability, satisfactory finite-sample performance, and computational efficiency.

The model is estimated using a sample of 1969 couples from 19 countries in the Survey of Health, Ageing, and Retirement in Europe (SHARE) data. Results provide empirical evidence of the existence of complementarities in the retirement process among European couples. A simulation exercise shows that complementarity reduces the median retirement age by approximately 5 months for wives and 2 months for husbands.

**Literature.** This paper broadly relates to two strands of literature. The first is the literature on choice-model estimation methods, including the textbook by Pudney (1989), the recent survey by Blomquist et al. (2023), and simulation-based approaches such as Smith Jr (1993), Gouriéroux and Monfort (1997), Celeux et al. (1996), and Dempster et al. (1977). Our main contribution is to connect a broad class of optimization-based choice models with kinks, corners, or nonconvexities—arising either from nonlinear budget constraints or from nonsmooth operators in the objective function—to censored latent-variable models. The equivalence clarifies how these models' one-to-many mapping from observed choices to unobservables parallels the structure of censoring, thereby enabling estimation through incomplete-data methods such as the stochastic EM algorithm.<sup>1</sup>

The paper also contributes empirically to the retirement literature, particularly studies considering the role of spousal interactions in retirement decisions, by providing empirical

<sup>1</sup>Bertanha et al. (2023) discuss and use the connection between bunching behavior and a censored model for estimation; we generalize and formalize this connection in a broader setup.

evidence of spousal complementarities in retirement decisions among European couples (Blundell et al., 2016; García-Miralles and Leganza, 2024; Honoré and de Paula, 2014, 2018; Hospido and Zamorro, 2014; Johnsen et al., 2022; Lalive and Parrotta, 2017; Michaud and Vermeulen, 2011; Michaud et al., 2020).<sup>2</sup>

The rest of the paper proceeds as follows. Section 2 characterizes a class of choice models with piecewise-smooth objective functions, develops the transformation of the original optimization problem into a potential-outcome framework, and introduces the SEM estimation method. Section 3 presents two illustrative examples. Section 4 applies the method to the interdependent duration model to empirically analyze joint retirement behavior among European couples. Finally, Section 5 concludes.

## 2 | CHOICE MODELS WITH CENSORED REDUCED FORM

In this section, we characterize a class of optimization-based choice models with piecewise smooth objective functions, which have a reduced form resembling a censored model, providing an alternative solution approach and facilitating the use of the stochastic Expectation-Maximization (SEM) estimation method for these choice models.

**Choice models.** We start with choice models that characterize units' behavior. For each unit  $i = 1, \dots, N$ , a *continuous* decision  $y_i$  is made by maximizing an objective function  $V(y; x_i, u_i, \theta)$ :

$$y_i = \arg \max_{y \in S_y} V(y; x_i, u_i, \theta). \quad (1)$$

For example,  $V(y)$  may represent an individual's utility function or a firm's profit function. While  $x_i$  and  $y_i$  are observable to researchers,  $u_i$  is unobservable and is assumed to follow a distribution  $f_u(u; \gamma)$ . The unknown parameter vector  $\kappa \equiv [\theta; \gamma]$  is of interest.<sup>3</sup>

The estimation of such models can be straightforward when the objective function  $V(y)$  is well behaved such that policy functions characterizing the optimal choice  $y_i$  can be derived. In such cases, one could pursue likelihood-based estimation, since it is fully specified, or exploit the moment restrictions implied by the model.

However, we focus on a specific class of models in which  $V(y)$  exhibits non-differentiability due to its piecewise feature, making the solution to the choice model problem and thus the estimation procedure more complex. Specifically,

$$V(y; x_i, u_i, \theta) = \sum_{k=1}^K \mathbb{1}(y \in S_k) V_k(y; x_i, u_i, \theta), \quad (2)$$

where  $\{S_k\}_{k=1}^K$  are mutually exclusive and collectively exhaustive subsets of  $S_y$ , meaning  $\bigcup_{k=1}^K S_k = S_y$  and  $S_k \cap S_j = \emptyset$  for  $k \neq j$ . The model is piecewise in the sense that the functional form of the objective depends on the region in which the action  $y$  lies.

This piecewise structure can arise from many sources in different applications, such as nonlinear budget constraints (e.g., progressive taxation) and nonlinear components in the objective function (e.g., max and min operators), as discussed later in Sections 3–4.

<sup>2</sup>See also Hurd (1990), Blau (1998), Gustman and Steinmeier (2000), Coile (2003), Gustman and Steinmeier (2004), Blau and Gilleskie (2006), Banks et al. (2007), Van der Klaauw and Wolpin (2008), Pozzoli and Ranzani (2009).

<sup>3</sup>The framework can be extended to incorporate unknown unit-specific heterogeneity with panel data, and the following discussion still applies.

This structure introduces two main challenges: (i) non-differentiability at the boundaries between regions  $S_k$ , which complicates the model solution; and (ii) solutions  $y_i$  at these boundaries often lead to censoring, meaning that multiple values of  $u_i$  are consistent with the observed  $y_i$ .

A common approach to estimating such models is to sequentially obtain local maximizers in each region  $S_k$  and then select the global maximizer, paired with simulation-based estimation methods such as indirect inference (Smith Jr, 1993; Gouriéroux and Monfort, 1997). This approach, however, requires additional care to account for possible corner solutions within each region, involves specifying an auxiliary model, and is generally less statistically efficient than MLE as well as less transparent in terms of identification. From a computational perspective, the simulated moments may be non-smooth, and the method can be computationally demanding in large parameter spaces.<sup>4</sup>

In the remainder of this section, we propose a new approach to estimating piecewise choice models. We show that under certain assumptions, the choice model can be transformed into a potential outcome framework, which we then estimate using the SEM procedure. This method avoids the need for region-by-region optimization, operates directly on the likelihood or moment conditions implied by the original model, and offers potential gains in computational efficiency, thus presenting a transparent and practically effective alternative.

## 2.1 | Potential Outcome Framework

Now we characterize a class of choice models whose optimization problem can be reformulated as a latent "potential outcome" framework, where the potential outcomes correspond to the maximizers of functions  $V_k(y)$ , both global and boundary-local. This framework will be further leveraged for estimation in the following subsection.

We impose the following assumptions. Assumption 1 is a regularity condition ensuring well-defined solutions, while Assumptions 2–3 are structural conditions on the objective function and its domain.

**Assumption 1 (Existence and uniqueness of the solution)** *For any  $\theta \in \Theta$ , the objective function  $V(y)$  attains a unique maximizer a.s. with respect to  $(x_i, u_i)$  over the domain  $S_y$ .*

**Assumption 2 (Global unimodality of auxiliary functions)** *For all  $(x_i, u_i)$  and  $\theta \in \Theta$ , and for any region  $S_k$ ,  $k = 1, \dots, K$ , that contains nonempty interior, there exists an open set  $\tilde{S}_k \supseteq S_k$  such that  $V_k(y)$  is continuous on  $\tilde{S}_k$ , admits no local maxima other than a unique global maximum, and the maximizer  $y_k^* \in \tilde{S}_k$  is characterized by the condition  $g_k(y_k^*) = 0$ .*

Assumption 2 imposes assumptions on each auxiliary function  $V_k(y)$ : while the overall objective function  $V(y)$  may exhibit irregular features globally, each component  $V_k(y)$  is assumed to be well-behaved, possessing a unique global maximizer on an open set  $\tilde{S}_k \supseteq S_k$  and no other local maxima. The domains  $\tilde{S}_k$  need not be identical across  $k$ , nor must they coincide with the global feasible set  $S_y$ . Moreover, when  $V_k(y)$  is differentiable, the condition  $g_k(y_k^*) = 0$  corresponds to the standard first-order condition. Assumption 2 ensures that the global maximizers of the auxiliary functions are well-defined.

Assumptions 1–2 allow for a certain degree of discontinuity. A special case is when  $V(y)$  exhibits jumps at region boundaries. For example, Section 3 presents an example with removable jumps due to fixed adjustment cost.

<sup>4</sup>While approaches such as Chernozhukov and Hong (2003) reformulate this problem as a quasi-posterior sampling task, they can require long sampling runs, especially when the parameter dimension is large.

**Assumption 3 (Existence of boundary maximizers)** For all  $(x_i, u_i)$  and  $\theta \in \Theta$ , and for any  $k \neq j$ , let  $S_{kj} \equiv \partial S_k \cap \partial S_j$  denote the common boundary between regions  $S_k$  and  $S_j$ , and let  $S_{k0} \equiv \partial S_k \cap \partial S_y$  denote the portion of the boundary of  $S_k$  that coincides with the boundary of the feasible set  $S_y$ . For any nonempty set  $S_{kj}$  with  $0 \leq j < k \leq K$ , the boundary maximizer  $y_{kj}^* \equiv \arg \max_{y \in S_{kj}} V(y)$  exists and is characterized by an equation  $g_{kj}(y_{kj}^*) = 0$ .

Assumption 3 ensures that we can characterize local maximizers of the objective function along all boundaries. Together with the global maximizers  $y_k^*$ , the boundary-local maximizers  $y_{kj}^*$  will serve as auxiliary latent variables, forming the basis of the potential outcome framework discussed in the remainder of this section.

The following result formalizes how these auxiliary variables characterize the global maximizer of the objective function.

**Theorem 1 (Latent potential outcome representation)** Under Assumptions 1-3, the global maximizer  $y$  of the objective function  $V(y)$  over the feasible set  $S_y$  must belong to the following finite candidate set:

$$y \in \left( \bigcup_{\{k: S_k^\circ \neq \emptyset\}} \{y_k^*\} \right) \cup \left( \bigcup_{0 \leq j < k \leq K} \{y_{kj}^*\} \right).$$

Moreover, the mapping from  $y$  to the elements of this set satisfies:

$$\begin{aligned} \text{If } y \in S_k^\circ, \text{ then } y_k^* &= y, \text{ for } 1 \leq k \leq K; \\ \text{If } y \in S_{kj}, \text{ then } y_{kj}^* &= y, \text{ for } 0 \leq j < k \leq K, \end{aligned}$$

where  $S_k^\circ$  denotes the interior of region  $S_k$ .

**Proof** By Assumption 1, a global maximizer  $y$  exists over the feasible set  $S_y$ . Since the sets  $\{S_k\}_{k=1}^K$  are mutually exclusive and collectively exhaustive,  $y$  must lie either in the interior of some region  $S_k^\circ$  for  $k = 1, \dots, K$ , or on a boundary  $S_{kj}$  for  $0 \leq j < k \leq K$ .

If  $y \in S_k^\circ$  for  $k = 1, \dots, K$ : On  $S_k^\circ$ ,  $V(y) = V_k(y)$  by construction, so  $y$  is also a local maximizer of  $V_k$ . Since  $V_k$  has no other local maxima except the unique global maximum  $y_k^*$  (by Assumption 2), it follows that  $y = y_k^*$ .

If  $y \in S_{kj}$  for  $0 \leq j < k \leq K$ : Since  $y$  maximizes  $V(y)$  globally and lies in  $S_{kj}$ , it must coincide with the boundary-local optimizer  $y_{kj}^*$ , by Assumption 3.

Therefore, the global maximizer must lie in the finite candidate set:

$$y \in \left( \bigcup_{k=1}^K \{y_k^*\} \right) \cup \left( \bigcup_{0 \leq j < k \leq K} \{y_{kj}^*\} \right).$$

First, Theorem 1 offers an alternative perspective on solving the optimization problem compared to the region-based optimization approach. By reframing the problem in this way, it avoids the need to discuss corner solutions that typically arise when searching for a local optimizer.

More importantly, by introducing the well-defined auxiliary latent variables  $y_k^*$  for  $k = 1, \dots, K$  and  $y_{kj}^*$  for  $k \neq j$ , Theorem 1 transforms the original optimization problem into

a *potential outcome* framework:

$$g_k(y_k^*; x, u, \theta) = 0, \text{ for } 1 \leq k \leq K; \quad g_{kj}(y_{kj}^*; x, u, \theta) = 0, \text{ for } 0 \leq j < k \leq K, \quad (3)$$

$$y = \sum_{k=1}^K \mathbb{1}(y \in S_k^0) y_k^* + \sum_{0 \leq j < k \leq K} \mathbb{1}(y \in S_{kj}) y_{kj}^*. \quad (4)$$

where  $u \sim f_u(u; \gamma)$ , and parameter vector  $\kappa \equiv [\theta; \gamma]$  is of interest.

Specifically, the auxiliary variables  $y_k^*$  and  $y_{kj}^*$ , defined by the functions  $g(\cdot)$ , can be interpreted as latent potential options available to unit  $i$ , as in Equation (3). Unit  $i$  selects the option that maximizes the objective function  $V(y)$ , and only this chosen outcome is observed by the researcher, as formalized in Equation (4). It is important to emphasize that unit  $i$  optimizes over the entire feasible domain  $S_y$ ; the finite set of alternatives in Equations (3)–(4), comprising  $y_k^*$  and  $y_{kj}^*$ , merely provides an equivalent representation that captures all possible optima and serves as a practical device for estimation.

#### Censored models.

Equations (3)–(4) constitute a valid framework, equivalent to the choice model in the likelihood as long as Assumptions 1–3 are satisfied. In simpler cases, where there is a one-to-one mapping between each observed choice  $y$  and the unobservable  $u$  through the functions  $g(\cdot)$ , direct estimation based on Equations (3)–(4) is often straightforward. Our main interest, however, lies in cases where some observed choices  $y$  are consistent with a range of unobservable values, creating a one-to-many mapping between  $y$  and  $u$ . In this case, the framework resembles a censored model, and its estimation—such as via Maximum Likelihood—often requires integrating over the entire feasible range of unobservables for each given choice  $y$ , making computation more demanding.

In the following subsection, we propose a simulation-based estimation method, the SEM method, to estimate the choice model leveraging the potential outcome framework.<sup>5</sup> Our assumptions are that the choice model is identified under the regularity conditions following Newey and McFadden (1994), ensuring that the targeted estimator (either MLE or GMM) is consistent and asymptotically normal. The focus on this paper is on the estimation procedure of the model.

## 2.2 | SEM Estimation Method

The stochastic EM (SEM) algorithm (Celeux et al., 1996) is a simulated version of the classical Expectation-Maximization (EM) algorithm (Dempster et al., 1977). As an iterative method, SEM exchanges between an E-step, where latent variables are sampled from their posterior distribution conditional on the observables, and an M-step, where we estimate the model using data and latent draws, until the parameters converge to the stationary distribution.

Based on the potential outcome framework in Equations (3)–(4), we propose the following general steps of conducting SEM. Define the latent variable vector  $y^* = [y_1^*, \dots, y_K^*, y_{1,2}^*, \dots, y_{K-1,K}^*]$ . Starting from an initial guess  $\hat{\kappa}^{(0)}$ , we iterate over the following two steps for  $s = 1, 2, \dots, S$  until convergence to a stationary distribution:

\* E-step: Given  $\hat{\kappa}^{(s)}$ , draw  $y_i^*$  from the posterior distribution  $f(y_i^* | y_i, x_i; \hat{\kappa}^{(s)})$ .

<sup>5</sup>Alternatively, the framework can also be readily utilized for Bayesian estimation methods, especially given its similarity to SEM in the sampling of latent variables. However, we will not pursue this further.

\* M-step: Estimate the model using the E-step draws :

$$\hat{\kappa}^{(s+1)} = \arg \min_{\kappa} H(\kappa; y^*, x)^T \times W \times H(\kappa; y^*, x),$$

where the known function  $H(\cdot)$  can be either the score functions or the moment restrictions based on functions  $g(\cdot)$ , and  $W$  is a weighting matrix.

The final estimator is the average of the last  $S_0$  iterations,  $\hat{\kappa} = \sum_{s=S-S_0+1}^S \hat{\kappa}^{(s)} / S_0$ .

The E-step requires drawing values for  $y_k^*$  and  $y_{kj}^*$  given the observables at parameter value  $\hat{\kappa}^{(s)}$ . In practice, for each observation  $i$ , if Equations (3)–(4) imply a unique value of  $u_i$  consistent with  $y_i$ , we recover  $u_i$  and compute  $y_k^*$  and  $y_{kj}^*$ . If these equations imply a range of values of  $u_i$  consistent with  $y_i$ , we sample  $u_i$  from the truncated distribution of  $f_u(u; \hat{\gamma}^{(s)})$  restricted to that range and then compute  $y_k^*$  and  $y_{kj}^*$ .

The M-step estimates the model in Equations (3)–(4) using pseudo-complete data. It can be either likelihood-based, where  $H(\cdot)$  denotes the score function, or moment-based, where  $H(\cdot)$  denotes moment conditions implied by the functions  $g(\cdot)$ . While a likelihood-based M-step yields an estimator asymptotically equivalent to the MLE under the regularity conditions in Nielsen (2000), a moment-based M-step yields an estimator defined by the specified moment conditions, often offering easier implementation and lower computational burden (Arcidiacono and Jones, 2003; Arellano and Bonhomme, 2016).

The advantage of SEM is that it avoids complex optimization involving integration over latent variables and instead consists of a sequence of much simpler M-step estimations under pseudo-complete data. While its baseline convergence speed may vary depending on the model, it can be improved, sometimes dramatically from hours to a few minutes, when combined with techniques such as parameter expansion (i.e., PX-SEM), as discussed in Wei (2024).

#### Difference from sequential optimization.

In principle, the sequential optimization can also be rewritten as an artificial potential outcome framework, albeit with a different set of potential outcomes. Specifically, let  $y_k^\dagger$  denote local optimizers within each region  $S_k$ . The following equations are then satisfied:

$$g_k^\dagger(y_k^\dagger; x, u, \theta) = 0, \quad \text{for } 1 \leq k \leq K, \quad \text{and } y = \sum_{k=1}^K \mathbb{1}(y \in S_k) y_k^\dagger.$$

While this shares the same structural form as our potential outcome framework, it is not as appealing because each function  $g_k^\dagger$ , as it comes from local optimization, corresponds to a truncated model (e.g., a Tobit mapping) and inherits the non-smoothness at region boundaries that characterizes the original choice model. In contrast, our approach introduces latent variables  $y_k^*$  such that the system  $g_k$  is smooth and globally valid. Thus, although the sequential method can be cast in the same formal structure, it does not offer the computational advantages achieved by our method.

#### Potential extensions and limitations.

This framework has the potential to be extended to more complex settings, such as panel data models with unobserved heterogeneity. In such cases, the E-step would require drawing values for both the introduced potential outcomes and the unobserved heterogeneities. It also accommodates cases where  $S_k$  is unit-specific, as demonstrated in the  $(S, s)$  model case in Section 3. The framework can also be extended to cases where  $y$  is a vector, as illustrated in the empirical application in Section 4 where  $y$  is a vector with two entries.

However, whether this solution and estimation method can be applied to a given choice model within our framework depends on the feasibility of determining the posterior distribution of  $y^*$  given  $y$  and other observables (i.e., the E-step). For example, when  $y$  is high dimensional—such as in dynamic models involving decisions in multiple periods—or when there are too many potential outcomes  $y^*$ , identifying the specific set of  $u$  values that are consistent with the observed choice can become very complex. While our method remains theoretically applicable, it may not be practically attractive in such cases. Therefore, to what extent our method can be applied to complicated dynamic models remains an open question.

### 3 | ILLUSTRATIVE EXAMPLES

In this section, we present illustrative examples that fit within our framework.

#### 3.1 | Nonlinear Budget Constraint

First, we consider an example where the piecewise features in the objective function arise from the nonlinearity of the budget set. Specifically, we illustrate our solution and estimation methods in the context of responses to progressive income taxation, which induces a piecewise-linear budget constraint, following (Saez, 2010).

In the following standard static model, agents choose their working hours to maximize utility from consumption (after-tax income) and leisure, subject to a progressive income tax. The utility function  $U(c, y; u, \theta) = c - (u/(1 + 1/\theta)) (y/u)^{(1+1/\theta)}$  depends positively on after-tax income  $c$  and negatively on before-tax income  $y$  (and thus positively on leisure), and  $\theta$  captures the compensated elasticities. Here,  $u > 0$  represents individual-specific ability, distributed according to  $f_u(u; \gamma)$ , and  $\kappa \equiv [\theta; \gamma]$  is the unknown parameter vector. The progressive income tax introduces a kink in the budget set, defined as  $c = (1 - t_0)y \mathbb{1}(y \leq \tau) + ((t_1 - t_0)\tau + (1 - t_1)y) \mathbb{1}(y > \tau)$ , where  $t_1 > t_0$ , indicating that income below  $\tau$  is taxed at rate  $t_0$ , while income above  $\tau$  is taxed at the higher rate  $t_1$ .

The kink at  $y = \tau$  naturally creates bunching behavior, and the size of the bunching mass can be exploited to recover elasticity  $\theta$  (Saez, 2010). While widely used in applied work, recent literature has pointed out the lack of nonparametric identification from the bunching mass alone, and the implicit assumptions underlying bunching estimators (Blomquist et al., 2015, 2021; Bertanha et al., 2023). In what follows, we instead rely on a fully specified structural model and estimate the underlying parameters using the model-implied likelihood, applying our solution and estimation methods.

It is straightforward to show that the optimal labor supply  $y$  results from the following optimization problem with a piecewise smooth objective function, satisfying our Assumptions 1-3:

$$y = \arg \max_{y > 0} V_1(y; u, \theta) \mathbb{1}(y \leq \tau) + V_2(y; u, \theta) \mathbb{1}(y > \tau), \quad (5)$$

where  $V_1 = U((1 - t_0)y, y; u, \theta)$  and  $V_2 = U((t_1 - t_0)\tau + (1 - t_1)y, y; u, \theta)$ . With auxiliary variables  $y_1^*$ ,  $y_2^*$ , and  $y_{1,2}^*$  defined as  $y_1^* = (1 - t_0)^\theta u$ ,  $y_2^* = (1 - t_1)^\theta u$ ,  $y_{1,2}^* = \tau$ , and the sets and subsets defined as  $S = S_y = \{y | y > 0\}$ ,  $S_1 = \{y | 0 < y \leq \tau\}$ ,  $S_2 = \{y | y > \tau\}$ , Theorem 1 implies that the original optimization problem in Equation (5) can be transformed into the

following potential outcome framework:

$$\begin{aligned} y_1^* &= (1 - t_0)^\theta u, \quad y_2^* = (1 - t_1)^\theta u, \quad y_{1,2}^* = \tau \\ y &= y_1^* \mathbb{1}(y < \tau) + y_2^* \mathbb{1}(y > \tau) + y_{1,2}^* \mathbb{1}(y = \tau) \end{aligned} \quad (6)$$

where  $u \sim f_u(u; \gamma)$ .

We then apply SEM for the estimation. Given the parameter guess  $\hat{\kappa}^{(0)}$ , the estimation proceeds as follows:

\* E-step: Draw values for  $y_1^*$  and  $y_2^*$  from the posterior distribution given  $y$

- For  $y \in S_1^o$ , we have  $y_1^* = y$ ; solve  $u$  from  $y_1^* = (1 - t_0)^{\hat{\theta}^{(s)}} u = y$  to compute  $y_2^*$
- For  $y \in S_2^o$ , we have  $y_2^* = y$ ; solve  $u$  from  $y_2^* = (1 - t_1)^{\hat{\theta}^{(s)}} u = y$  to compute  $y_1^*$
- For  $y = \tau$ , define  $S_u \equiv \{u | \tau / (1 - t_0)^{\hat{\theta}^{(s)}} \leq u \leq \tau / (1 - t_1)^{\hat{\theta}^{(s)}}\}$  where  $y_2^* \leq \tau \leq y_1^*$  leading to a corner solution, draw  $u$  from  $f_u(u | u \in S_u; \hat{\gamma}^{(s)})$  to compute  $(y_1^*, y_2^*)$

\* M-step: Update to  $\hat{\kappa}^{(s+1)}$  using  $y_1^*$  and  $y_2^*$  based on equation sets (6).

The E-step and M-step are iterated until  $\hat{\kappa}^{(s)}$  converges to a stationary distribution. The final  $S_0$  iterations of  $\hat{\kappa}^{(s)}$  are averaged to obtain the final estimator. Appendix A presents two simulation exercises—one with two kinks and one with covariates—along with their estimation results from the moment-based M-step.<sup>6</sup>

### 3.2 | (S, s) Model

Another example of piecewise structure is provided by models with non-smooth cost functions, such as piecewise linear costs from block pricing. In this subsection, we focus on a specific case with a fixed adjustment cost, which leads to the well-known (S, s) rule.<sup>7</sup> We use a simple example to illustrate the microfoundation and show that introducing a fixed adjustment cost makes the objective a special case of our choice-model framework. We then propose the SEM algorithm to estimate the resulting (S, s) rule.

Consider a simplified static stock adjustment problem in the context of durable goods:

$$V(y; y_1, u, \theta) = U(y; u, \theta) - K \cdot \mathbb{1}\{y \neq y_1\}$$

where  $y_1$  denotes the stock value at the beginning of the period,  $U(\cdot)$  is the indirect utility function, and  $K > 0$  is a fixed adjustment cost. The objective is a special case of our framework, with a removable discontinuity at  $y = y_1$ .<sup>8</sup> If  $U(\cdot)$  is concave and Assumptions 1–3 hold, we can construct two unique potential outcomes:  $y_{1,2}^* = y_1$  and  $y_1^* = y^* \equiv \arg \max_y U(y) - K$ , of which only one is realized. It is easy to show that the result delivers the (S, s) rule characterized by a lower bound, an upper bound, and the target value: only when  $y_1$  is below the lower bound or above the upper bound is the adjustment made to the

<sup>6</sup>The connection between bunching behavior and a censored model has also been noted by Bertanha et al. (2023) and exploited to propose estimators under different assumptions on  $f_u(u; \gamma)$ . Our estimation method is complementary and can be readily applied to their setting.

<sup>7</sup>Originally introduced in the inventory literature (see Arrow et al. 1951; Scarf 1960), the (S, s) model has also been applied to study other infrequent adjustments, such as durable goods (Grossman and Laroque 1990; Caballero 1993; Eberly 1994; Attanasio 2000; Bover 2010; Attanasio et al. 2022; Richard 2024, among others) and firms' pricing decisions (e.g., in a recent application by Karadi et al. 2025).

<sup>8</sup>Here  $S_1 = \{y < y_1\}$ ,  $S_3 = \{y > y_1\}$ ,  $S_2 = S_{1,2} = S_{2,3} = \{y_1\}$ ,  $V_1(y) = V_3(y) = U(y; x, u, \theta) - K$ , and  $V_2(y) = U(y; x, u, \theta)$ . Note that in this case, the region definitions are individual specific.

target  $y^*$ .<sup>9</sup>

In the remainder of the subsection, we construct an empirical framework of the  $(S, s)$  model for durable goods adjustment. Following [Attanasio \(2000\)](#) and [Bover \(2010\)](#), we characterize the lower and upper bounds and the target value flexibly as functions of observables  $x$  and  $z$  and unobservables  $u$  and  $v$ , for which we then propose the SEM estimation method. Let  $y^*$  denote the latent target stock value,  $y$  the realized stock value, and  $b^u$  and  $b^l$  the unobserved upper and lower thresholds, respectively. The model is specified as:

$$\begin{aligned} y^* &= x'\beta + \sigma u, & b^u &= y^* + \exp(z'\gamma_u + \omega v), & b^l &= y^* - \exp(z'\gamma_l + \omega v), \\ y &= \tilde{y}_1 \mathbb{1}(b^l < y_1 < b^u) + y^* \mathbb{1}(y_1 \geq b^u \text{ or } y_1 \leq b^l). \end{aligned}$$

where  $(u, v)$  are jointly standard normal with correlation coefficient  $\rho$ . The unknown parameter vector is  $\kappa \equiv [\beta; \gamma_u; \gamma_l; \sigma; \omega; \rho]$ .

As the equations indicate, adjustment occurs only when  $y_1$  crosses the thresholds  $b^l$  or  $b^u$ , triggering an upward or downward shift to the target level  $y^*$ ; otherwise, the stock value evolves to  $\tilde{y}_1$ , whose dynamics are assumed to be exogenous and of less interest. The adjustment action is observed and captured by the variable  $d$ :  $d = 1$  represents an upgrade,  $d = -1$  a downgrade, and  $d = 0$  indicates no adjustment (although the value may still appreciate or depreciate to  $\tilde{y}_1$ ).<sup>10</sup>

So far we have shown that the  $(S, s)$  durable goods model is a natural illustration of our framework: the fixed adjustment cost makes the objective piecewise smooth, and the reduced form fits into a potential outcome framework. Thus, it provides a convenient setting to apply the SEM method. Starting from an initial parameter guess  $\hat{\kappa}^{(s)} = [\hat{\beta}^{(s)}; \hat{\gamma}_u^{(s)}; \hat{\gamma}_l^{(s)}; \hat{\sigma}^{(s)}; \hat{\omega}^{(s)}; \hat{\rho}^{(s)}]$ , we iterate the following E-step and M-step until the convergence of  $\hat{\kappa}^{(s)}$  to a stationary distribution:

\* E-step: Draw values for  $y^*$  from the posterior distribution given  $y$

- For  $d = 1$  and  $d = -1$ , set  $y^* = y$
- For  $d = 0$ , define  $S_{uv} \equiv \{(u, v) | y_1 - \exp(z'\gamma_u^{(s)} + \omega^{(s)}v) < x'\beta^{(s)} + \sigma^{(s)}u < y_1 + \exp(z'\gamma_l^{(s)} + \omega^{(s)}v)\}$ , draw  $(u, v)$  from  $\Phi(u, v | (u, v) \in S_{uv})$ , and compute  $y^*$

\* M-step: Update to  $\hat{\kappa}^{(s+1)}$  using  $y^*$ . For simplicity, we propose the following moment-based sequential estimation:

- Regress  $y^*$  on  $x$ , obtain  $\hat{\beta}^{(s+1)}$  and  $\hat{\sigma}^{(s+1)}$ , calculate residuals  $\hat{u}$
- Probit regression to obtain  $\hat{\gamma}_l^{(s+1)}$ ,  $\hat{\gamma}_u^{(s+1)}$ ,  $\hat{\rho}^{(s+1)}$ , and  $\hat{\omega}^{(s+1)}$  leveraging the following equation

$$\begin{aligned} \left( \mathbb{1}(d = -1 \text{ or } d = 1) | y^*, y_1, z, u \right) &= \Phi \left( \frac{1}{\omega \sqrt{1 - \rho^2}} \log(|y_1 - y^*|) \right. \\ &\quad \left. - \frac{1}{\omega \sqrt{1 - \rho^2}} z'\gamma_l \mathbb{1}(y^* > y_1) - \frac{1}{\omega \sqrt{1 - \rho^2}} z'\gamma_u \mathbb{1}(y^* < y_1) - \frac{\rho}{\sqrt{1 - \rho^2}} u \right) \end{aligned}$$

<sup>9</sup>The  $(S, s)$  rule has been shown to be an optimal policy under more general setups (see [Scarf 1960](#); [Grossman and Laroque 1990](#); [Caballero and Engel 1999](#)).

<sup>10</sup>In this example, the regions  $S_k$  are individual-specific, and the stock value without adjustment  $\tilde{y}_1$  can vary over time. Hence  $d$  is required to identify which potential outcome is realized.

After convergence to the stationary distribution, we take the average of the last  $S_0$  iterations of  $\hat{\kappa}^{(s)}$  as the final estimate. Simulation results are presented in Appendix A.

#### 4 | COMPLEMENTARITIES IN RETIREMENT DECISIONS AMONG EUROPEAN COUPLES

In this section, we apply our methods to a nontrivial example: the interdependent duration model proposed by [Honoré and de Paula \(2018\)](#) (hereafter, HDP). The HDP model can be viewed as a bivariate extension of the generalized accelerated failure time framework, accommodating simultaneity in decision-making, with the observed outcome interpreted as the Nash bargaining solution. The resulting objective function has a piecewise structure arising from the nonlinear  $\max(\cdot)$  operator.

While HDP examine joint retirement decisions among U.S. couples using the Health and Retirement Study (HRS), we instead investigate the joint retirement behavior of European couples using data from the Survey of Health, Ageing and Retirement in Europe (SHARE).<sup>11,12</sup> Our analysis builds directly on the HDP model but implements the transformation and estimation methods developed in Section 2.

In the remainder of this section, we first introduce the HDP model, discussing its piecewise features and the transformation to the potential outcome framework; we then present the SEM estimation method and the data, followed by the estimation results.

##### 4.1 | The HDP Model

Let  $t$  denote the family age. At time  $t = 0$ , a household consisting of a wife and a husband (denoted  $w$  and  $h$ ) determines their optimal retirement times  $T_i$ , for  $i \in \{w, h\}$ . Each member  $i$  receives an individual constant utility flow  $k_i > 0$  before retirement (for  $t < T_i$ ), reflecting labor market attachment. Correlation between  $k_w$  and  $k_h$  is allowed to capture potential assortative matching in marriage.

After retirement (for  $t > T_i$ ), member  $i$  receives a partner-dependent utility flow  $H_i(t)D(t, T_{-i})$ , where  $-i$  denotes the partner. Here,  $H_i(t) = Z_i(t)\varphi_i(x)$  represents baseline utility as a deterministic function of family age  $t$  and individual characteristics  $x_i$ .<sup>13</sup> We assume  $H_i(0) = 0$  and that  $H_i(t)$  is continuous and strictly increasing in  $t$ , and that  $\lim_{t \rightarrow \infty} H_i(t) = \infty$ , reflecting that retirement becomes more attractive over time and is treated as an absorbing state. The partner's retirement  $T_{-i}$  affects the utility flow through the scaling factor,  $D(t, T_{-i}) = (\delta - 1)\mathbf{1}(t \geq T_{-i}) + 1$ , with  $\delta \geq 1$ . When  $\delta > 1$ , there is utility complementarity, scaling the utility flow after retirement by  $\delta$  whenever both partners are retired.

##### Utility function.

Under this setup, the discounted utility of member  $i$ , if retiring at  $t_i$  given their partner's retirement time  $t_{-i}$ , can be expressed as:

$$U(t_i; x_i, k_i, t_{-i}) = \int_0^{t_i} k_i e^{-\rho s} ds + \int_{t_i}^{\infty} H_i(s) [(\delta - 1)\mathbf{1}(s \geq t_{-i}) + 1] e^{-\rho s} ds$$

<sup>11</sup>See, e.g., [Hospido and Zamarro 2014](#); [Johnsen and Willén 2022](#); [García-Miralles and Leganza 2024](#); [Michaud et al. 2020](#) on joint retirement and leisure complementarities.

<sup>12</sup>[Honoré and de Paula \(2014\)](#) apply the HDP model to SHARE. Our analysis uses updated data and a different estimation procedure.

<sup>13</sup>Although  $H_i(t; x_i)$  depends on  $x_i$ , we suppress it in the notation.

where  $\rho$  is the discount rate.

#### Nash Bargaining.

Given realizations of  $(k_w, k_h)$ , the couple chooses retirement times through Nash bargaining, a common approach in the intra-household literature (Chiappori et al., 2012):

$$(T_w, T_h) = \arg \max_{t_w, t_h} (U(t_w; x_w, k_w, t_h) - A_w) \times (U(t_h; x_h, k_h, t_w) - A_h) \quad (7)$$

where  $A_i$  denotes the threat point, specified as a fixed fraction of the maximum utility that individual  $i$  would achieve in the absence of complementarity, following HDP.

If  $\delta = 1$ , i.e. no utility complementarities, optimal retirement occurs when  $k_i = H_i(T_i)$ , that is, when  $Z_i(T_i)\varphi_i(x) = k_i$ . Assuming  $k_i$  follows an exponential distribution with mean 1,  $Z_i(t) = t^{\alpha_i}$ , and  $\varphi_i(x) = \exp(x'\beta_i)$ , the solution reduces to a Weibull proportional hazards model. By contrast, if  $\delta > 1$ , an individual may deviate from the above condition to benefit from complementarities, generating joint retirement patterns.

## 4.2 | Piecewise Structure and the Potential Outcome Framework

We now show that the objective function of the Nash bargaining problem in Equation (7), denoted by  $V(t_w, t_h)$ , is piecewise smooth and fits into our general framework.

Define  $\tilde{H}_i(t) \equiv \int_t^\infty H_i(s)e^{-\rho s} ds$  and  $G_i(t) \equiv \int_0^t k_i e^{-\rho s} ds + \int_t^\infty H_i(s)e^{-\rho s} ds = k_i \rho^{-1}(1 - e^{-\rho t}) + \tilde{H}_i(t)$ . Then Equation (7) can be rewritten as:

$$(T_w, T_h) = \arg \max_{(t_w, t_h)} \left( G_w(t_w) + (\delta - 1)\tilde{H}_w(\max(t_w, t_h)) - A_w \right) \times \left( G_h(t_h) + (\delta - 1)\tilde{H}_h(\max(t_w, t_h)) - A_h \right).$$

The presence of the nonlinear  $\max(\cdot)$  operator makes the objective function piecewise defined. Specifically, depending on whether  $t_w > t_h$  or  $t_w < t_h$ , the functional form of  $V(t_w, t_h)$  differs:

$$V(t_w, t_h) = \mathbb{1}(t_w \geq t_h)V_w(t_w, t_h) + \mathbb{1}(t_w < t_h)V_h(t_w, t_h), \quad (8)$$

where

$$V_w(t_w, t_h) = \left( G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w \right) \times \left( G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h \right),$$

$$V_h(t_w, t_h) = \left( G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w \right) \times \left( G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h \right),$$

with the objective function on the common boundary  $t_w = t_h$  given by:

$$V_j(t, t) \equiv \left( G_w(t) + (\delta - 1)\tilde{H}_w(t) - A_w \right) \left( G_h(t) + (\delta - 1)\tilde{H}_h(t) - A_h \right).$$

Appendix B verifies that this model satisfies Assumptions 1-3. In particular, it shows that the auxiliary functions  $V_w(t_w, t_h)$  and  $V_h(t_w, t_h)$  possess no local maximizers other than unique global ones, which are further characterized by the first-order conditions. Specifically, let  $op_w \equiv (T_w^w, T_h^w)$  and  $op_h \equiv (T_w^h, T_h^h)$  denote the optimizers of  $V_w(t_w, t_h)$  and  $V_h(t_w, t_h)$ , respectively; that is,  $op_w = \arg \max_{(t_w, t_h)} V_w(t_w, t_h)$ ,  $op_h = \arg \max_{(t_w, t_h)} V_h(t_w, t_h)$ . Then the conditions that  $op_w$  and  $op_h$  must satisfy are given below, with the detailed deriva-

tions provided in Appendix B:

$$g_w(\text{op}_w) \equiv \begin{bmatrix} T_h^w - H_h^{-1}(k_h) \\ \psi(T_w^w, T_h^w, x_w, x_h, k_w, k_h) \end{bmatrix} = \mathbf{0}, \quad (9)$$

$$g_h(\text{op}_h) \equiv \begin{bmatrix} T_w^h - H_w^{-1}(k_w) \\ \psi(T_h^h, T_w^h, x_h, x_w, k_h, k_w) \end{bmatrix} = \mathbf{0}, \quad (10)$$

where  $\psi(t_w, t_h, x_w, x_h, k_w, k_h) \equiv H_h(t_w)e^{-\rho t_w}(1-\delta) \times (G_w(t_w) + (\delta-1)\tilde{H}_w(t_w) - A_w) + (G_h(t_h) + (\delta-1)\tilde{H}_h(t_h) - A_h) \times (k_w e^{-\rho t_w} - H_w(t_w)\delta e^{-\rho t_w})$ .

Moreover, let  $\text{op}_j \equiv (T^j, T^j)$  denote the local optimizer along the boundary  $t_w = t_h$ . Appendix B characterizes the condition that  $\text{op}_j$  must satisfy:

$$g_j(\text{op}_j) \equiv \eta(T^j, T^j, x_w, x_h, k_w, k_h) = 0, \quad (11)$$

where  $\eta(t, t, x_w, x_h, k_w, k_h) = (k_w - \delta H_w(t))(G_h(t) + (\delta-1)\tilde{H}_h(t) - A_h) + (k_h - \delta H_h(t))(G_w(t) + (\delta-1)\tilde{H}_w(t) - A_w)$ .

**The potential outcome framework.**

Applying Theorem 1, we can now transform the original Nash bargaining problem into the following potential outcome framework:

$$\begin{aligned} g_w(\text{op}_w) = \mathbf{0}, \quad g_h(\text{op}_h) = \mathbf{0}, \quad g_j(\text{op}_j) = 0, \\ (T_w, T_h) = \text{op}_w \mathbb{1}(T_w > T_h) + \text{op}_h \mathbb{1}(T_w < T_h) + \text{op}_j \mathbb{1}(T_w = T_h), \end{aligned} \quad (12)$$

where the unknown parameters include the function  $H_i(\cdot)$ , the complementary parameter  $\delta$ , and the joint distribution of  $(k_w, k_h)$ .

As before, we can interpret  $\text{op}_w$ ,  $\text{op}_h$ , and  $\text{op}_j$  as three pairs of artificial potential choices available to each household, from which each household selects the optimal one among these finite options. For researchers, only the realized selection outcome is observed. This framework, with the introduced latent variables  $\text{op}_w$ ,  $\text{op}_h$ , and  $\text{op}_j$ , facilitates the SEM estimation method, which will be discussed in the following subsection.

As for the identification of the model, a detailed discussion and complete proof are provided by [Honoré and de Paula \(2018\)](#). The intuition behind the identification is as follows. For individuals whose partner's optimal retirement time is effectively infinite—due to very small values of  $\varphi(x)$ —their retirement decisions are approximately individual decisions and can thus be used to identify the function  $H_i(\cdot)$ . Conversely, for individuals whose partner's optimal retirement time is effectively zero—due to very large values of  $\varphi(x)$ —their retirement decisions help identify  $\delta$ . The identification of  $H_i(\cdot)$  and  $\delta$  then allows the recovery of the household-specific values of  $k_w$  and  $k_h$ , and thus the joint distribution of  $(k_w, k_h)$ .

### 4.3 | Estimation

This section explains the estimation process. We begin by describing the empirical specification and then discuss the estimation method.

### Empirical Specification.

We specify the functional form of  $H_i(t)$  by setting  $Z_i(t) = t^{\alpha_i}$  and  $\varphi_i(x) = \exp(x'\beta_i)$ , which gives  $H_i(t) = t^{\alpha_i} \exp(x'\beta_i)$ .<sup>14</sup> Following the HDP model, we allow for positive dependence between  $k_w$  and  $k_h$  via the Clayton copula. With Exp(1) margins for  $k_w$  and  $k_h$ , the joint survivor function is

$$S(k_w, k_h; \tau) = \begin{cases} \exp(-k_w) \exp(-k_h), & \text{if } \tau = 0, \\ (\exp(-k_w)^{-\tau} + \exp(-k_h)^{-\tau} - 1)^{-1/\tau}, & \text{if } \tau > 0, \end{cases} \quad (13)$$

where  $\tau = 0$  implies independence and  $\tau > 0$  induces positive dependence.

Although the model is fully parameterized, direct MLE is challenging because the likelihood function has no closed form and involves integrals arising from joint retirement and right-censoring. Specifically, for any given  $(T_w, T_h)$ , the solution for  $(k_w, k_h)$  obtained from systems (9–11) has no closed form because of the functions  $\psi(\cdot)$  and  $\eta(\cdot)$ . Moreover, under joint retirement ( $T_w = T_h$ ) or right-censoring, where only lower bounds of  $T_w$  and/or  $T_h$  are observed, the solution is no longer unique, with a set of potential  $(k_w, k_h)$  values consistent with  $(T_w, T_h)$ . Consequently, deriving the likelihood function requires integrating over all feasible values of  $(k_w, k_h)$ .

### The SEM Algorithm.

In the remaining subsection, we discuss the SEM algorithm adapted to this model, leveraging the potential outcomes framework (Equation system 12). Our E-step uses both direct sampling and MCMC, and the M-step maximizes moment conditions implied by the model, incorporating the parameter-expansion technique of Wei (2024) to accelerate convergence.

Let  $\kappa \equiv [\alpha_w \beta'_w \alpha_h \beta'_h \delta \tau]'$  denote the vector of unknown parameters. Starting from an initial guess  $\hat{\kappa}^{(0)}$ , we iterate between an E-step, where values of  $op_w, op_h$  are drawn from the posterior distribution implied by system (12) given the parameter guess  $\hat{\kappa}^{(s)}$ , and an M-step, where both the observables and the E-step draws are used to update the parameter guess to  $\hat{\kappa}^{(s+1)}$ , until the convergence to the stationary distribution.

**E-step.** The specific approach for sampling  $op_w, op_h$  depends on the retirement type and whether censoring is present. The following paragraphs briefly outline the sampling process, with detailed procedures provided in Appendix C.

- \* *Sequential retirement, no censoring* ( $T_w > T_h$ , or  $T_w < T_h$ ): Given the parameter guess  $\hat{\kappa}^{(s)}$ , the potential outcome framework in the system (12) identifies the *unique* pair of  $k_w$  and  $k_h$ , which can be used to compute values for all the latent variables.
- \* *Joint retirement, no censoring* ( $T_w = T_h$ ): Let  $T_w = T_h = t^*$ . System (12) implies  $op_j = (t^*, t^*)$ . With two unknowns,  $k_w$  and  $k_h$ , but only one condition,  $g_j(op_j; \hat{\kappa}^{(s)}) = 0$ , there exists a set of  $(k_w, k_h)$  satisfying the equation. However, natural bounds arise from the sequential-retirement limits, where one partner retires at  $T_i = t^*$  and the other at  $T_{-i} = t^* + \Delta$ , with  $\Delta \rightarrow +0$ . Solving  $g_w(t^* + \Delta, t^*; \hat{\kappa}^{(s)}) = 0$  and  $g_h(t^*, t^* + \Delta; \hat{\kappa}^{(s)}) = 0$  yields these bounds (Appendix C). Figure 1a illustrates the feasible set: the blue segment, possibly curved, represents all admissible  $(k_w, k_h)$ , with endpoints at the boundary cases. We then draw one  $(k_w, k_h)$  pair from this set for each household via MCMC, with probabilities proportional to the joint distribution implied by Equation (13), to compute  $op_w, op_h$ , and  $op_j$ .
- \* *Right censored*: When only the time of the couple's first retirement is observed, the case still corresponds to sequential retirement. For example, when we observe  $T_w = t_1^*$  and

<sup>14</sup>Unlike the HDP model, we simplify  $Z_i(t)$  by removing jumps at the retirement age. This simplification does not affect the estimation algorithm.

$T_h > t_2^* \geq t_1^*$ , System (12) then implies  $op_h = (T_w, T_h)$  and  $g_h(op_h; \hat{\kappa}^{(s)}) = \mathbf{0}$ . We can therefore solve for a unique  $k_w$  and a lower bound for  $k_h$ . By directly sampling  $k_h$  from the truncated conditional distribution given  $k_w$ , we can then compute the corresponding values of  $op_h$ ,  $op_w$ , and  $op_j$ .

When both retirement times are censored ( $T_w > t^*$ ,  $T_h > t^*$ ), the case may correspond to either sequential or joint retirement. Leveraging system (12), we solve for the feasible set of  $(k_w, k_h)$ . While detailed discussions are in Appendix C, Figure 1b depicts the feasible set: the blue area represents all possible  $(k_w, k_h)$ , with its boundary corresponding to the sequential retirement limits, where one partner retires at  $T_i = t^*$  and the other later at  $T_{-i} > t^*$  for  $i = \{w, h\}$ , as well as the joint retirement limit with  $T_w = T_h = t^*$ . In this case, we use rejection sampling: draw  $(k_w, k_h)$  from the joint distribution and accept the draw within the feasible set. Using these draws, we compute  $op_w$ ,  $op_h$ , and  $op_j$ .

**M-step.** We update  $\hat{\kappa}^{(s+1)}$  using both the data and the E-step draws of  $op_w$  and  $op_h$ . Specifically, we conduct sequential estimation to alleviate the computational burden.

- \*  $\hat{\alpha}_i^{(s+1)}$  and  $\hat{\beta}_i^{(s+1)}$ : Taking advantage of the fact that  $T_h^w$  in  $op_w$  and  $T_w^h$  in  $op_h$  satisfy  $T_h^w = H_h^{-1}(k_h)$  and  $T_w^h = H_w^{-1}(k_w)$ , respectively (systems 9–10), we estimate the following log-linear specification by OLS, since there is no censoring.<sup>15</sup>

$$\log(T_i^{-i}) = -\frac{1}{\alpha_i} x' \beta_i + \frac{1}{\alpha_i} \log(k_i) \text{ for } i \in \{w, h\}.$$

Here,  $(\log(k_i)|x) = -\gamma$  and  $(\log(k_i)|x) = \pi^2/6$ , with  $\gamma$  denoting the Euler-Mascheroni constant.

- \*  $\hat{\tau}^{(s+1)}$  and  $\hat{\delta}^{(s+1)}$ : We update the copula-dependence parameter  $\tau$  and the complementarity parameter  $\delta$  by matching the moment conditions implied by the joint distribution of  $(k_w, k_h)$ . While Appendix C details the moment-matching criterion and implementation, we outline the main steps here.

Given  $\hat{\alpha}_i^{(s+1)}$  and  $\hat{\beta}_i^{(s+1)}$ , for each candidate value of  $\delta$  we obtain the residual pair  $(k_w, k_h)$  from the equations  $g_w(op_w) = \mathbf{0}$ ,  $g_h(op_h) = \mathbf{0}$ , and  $g_j(op_j) = 0$ , and compute the implied Clayton-copula dependence parameter

$$\hat{\tau}(\delta) = \frac{2\kappa_{\text{endall}}(k_w(\delta), k_h(\delta))}{1 - \kappa_{\text{endall}}(k_w(\delta), k_h(\delta))},$$

where  $\kappa_{\text{endall}}$  denotes Kendall's rank correlation.

Since the joint moments of  $(k_w, k_h)$  under a non-zero Clayton copula lack closed-form expressions, we apply the Rosenblatt transformation (Rosenblatt, 1952) to map the dependent pair into independent variables  $(\tilde{k}_w(\hat{\tau}), \tilde{k}_h(\hat{\tau}))$ , whose moments are known. The values of  $(\hat{\delta}^{(s+1)}, \hat{\tau}^{(s+1)})$  are then chosen such that the empirical moments of  $(\tilde{k}_w, \tilde{k}_h)$  match their theoretical counterparts.

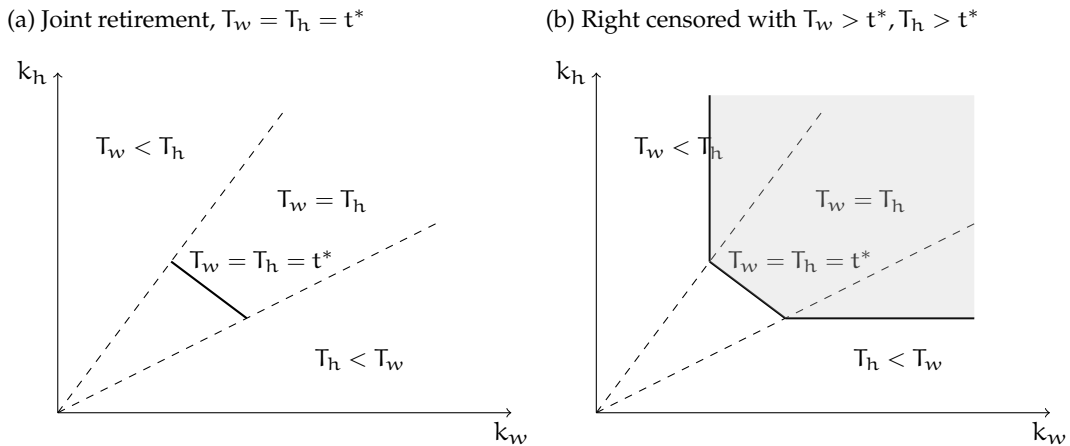
- \* *Extra step:* The baseline M-step described above yields a consistent and computationally efficient SEM estimator. In practice, however, when the sample is small or right-censoring is extensive, the M-step updates for  $\delta$  often exhibit high volatility across iterations, and the resulting SEM estimates display greater dispersion around the true parameter values in finite samples. To improve stability and precision in such cases, we augment the M-step with an additional unconditional moment condition that matches the model-implied share of joint retirements in the M-step to the empirical share implied by the E-step

<sup>15</sup>Alternatively, one could estimate two separate Weibull proportional-hazard models by maximum likelihood. We use the moment-based specification for simplicity and because it tends to be more stable and robust when the E-step draws deviate from the model assumptions; see Wei (2024) for discussion.

draws. This auxiliary restriction both sharpens the identification of  $\delta$  and regularizes its update, substantially improving the stability of the M-step and the precision of the SEM estimator without altering the asymptotic target. Although the additional step adds to the computation time relative to the baseline SEM, the overall algorithm remains highly efficient with multithreaded implementation.<sup>16</sup> Detailed procedures and simulation exercises are provided in Appendix C.

After convergence to the stationary distribution, the final estimate is obtained by averaging the last  $S_0$  iterations of  $\hat{\kappa}^{(s)}$ .

**FIGURE 1** Feasible sets for  $(k_w, k_h)$



2 Notes: The figures illustrate retirement decisions based on  $(k_w, k_h)$  values. The upper and lower sections, divided by the dashed line, represent regions of sequential retirement, while the middle shows joint retirement. The blue line in the left figure (possibly curved) indicates all possible  $(k_w, k_h)$  consistent with data. The right figure depicts the  $(k_w, k_h)$  region when both partners' retirement times are censored at  $t^*$ .

<sup>16</sup>Replicating HDP on HRS yields a similar estimate,  $\delta = 1.058$  (without jumps in  $Z_i(t)$  and with 42 coefficients). The SEM converges within 30 seconds, and 200 iterations take about 9 minutes, on par with the enhanced algorithm in Forneron (2023), who report that the original HDP procedure can take more than 5 hours to estimate 30 coefficients, whereas their improved method takes about 13 minutes.

#### 4.4 | Data

We use data from the Survey of Health, Ageing and Retirement in Europe (SHARE), a multidisciplinary biennial panel survey launched in 2004 that collects detailed information on the health, socioeconomic status, and social networks of individuals aged 50 and above across European countries and Israel. In the rest of this subsection, we describe the data-model variable mapping and sample selection.

The reference time corresponding to family age  $t = 0$  is set as twelve months before the first individual in the household reaches the early retirement age in their country. With this definition, we can then measure, for each member, the number of months after that reference point at which they retire or become right-censored, denoted  $t = T_i$ .<sup>17</sup> We use this mapping because retirement eligibility and timing vary substantially across European countries. In the model,  $t = 0$  corresponds to the moment when the household makes its retirement decision; fixing a common chronological age (e.g., 55) as the reference point would misalign the model with institutional realities in countries where retirement typically occurs later, such as around age 65.

Other control variables include the individual's age at period  $t = 0$ , the age difference between spouses (husband's age minus wife's age, in years), education level (primary, secondary, or tertiary, following the ISCED classification), a self-reported health indicator, the number of grandchildren, the number of children, household wealth, and regional dummies (Southern Europe, Eastern Europe, and Northern Europe).<sup>18,19</sup>

We restrict to couples in which both partners are in the labor force at  $t = 0$ .<sup>20</sup> We also exclude households in which any member was ever enrolled in a disability pension, as their decision-making process may differ, and households with missing information. Using data from 2006–2020 (waves 2–8, excluding wave 3), we obtain a sample of 1969 couples from 19 countries.<sup>21</sup> Appendix D presents summary statistics.

#### 4.5 | Results

Leveraging the potential outcome framework, we estimate the HDP model using the SEM algorithm. As shown in Appendix D, the M-step updates converge rapidly—within a few iterations and in about one to two minutes. In total, we run 200 iterations, which take roughly 15 minutes, and average the last 100 updates to obtain the final estimates reported in Table 1. Additional results on model fit are provided in Appendix D.

Turning to the estimation results, Table 1 shows that the shape parameters  $\alpha_w$  and  $\alpha_h$  are both greater than one, indicating that the exit rate accelerates with time. We also find substantial heterogeneity in retirement behavior across households.

Specifically, older households at the reference time tend to retire earlier. Consistent with [Honoré and de Paula \(2018\)](#), age differences increase the retirement hazard for husbands but

<sup>17</sup>Retirement is defined based on employment status: self-reported as "retired" and not working for pay. The retirement timing is determined using the retrospective retirement year and month.

<sup>18</sup>Household wealth is defined as the total value of bank accounts, government and corporate bonds, stocks, mutual funds, individual retirement accounts, and contractual savings for housing and life insurance, minus financial liabilities.

<sup>19</sup>For all time-varying variables, the value at  $t = 0$  is used. When the value is missing, we approximate with the value in the nearest two waves.

<sup>20</sup>This restriction substantially reduces the sample size and may make the sample less representative for some countries, as a large proportion of women either never worked or left the labor force before age 50 ([Hospido and Zamarro, 2014](#)).

<sup>21</sup>Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Italy, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Switzerland, Sweden. Wave 3 is excluded due to its different setup.

reduce it for wives. Education has a significant effect on husbands' retirement timing—more educated husbands retire later—while the effect is not significant for wives. Poor health has no significant effect for either spouse, likely because the sample excludes individuals who have ever received disability-related pensions, resulting in a relatively healthy sample. Having more grandchildren tends to accelerate wives' retirement but has no significant effect on husbands. We also observe substantial regional heterogeneity for both spouses.<sup>22</sup> Overall, the results are consistent with the HDP estimates using U.S. data.

The estimated complementarity parameter is  $\delta = 1.062$  (compared with 1.04-1.07 in HDP), implying that the utility flow from retirement increases by about 6% when one's partner retires. The copula dependence parameter  $\tau$  is approximately 0.7 (versus 0.5 in HDP), corresponding to a Kendall's rank correlation of roughly 0.26.<sup>23</sup>

---

<sup>22</sup>The interpretation of regional effects is less straightforward, since countries differ in the definition of the reference period ( $t = 0$ ) and in sample composition.

<sup>23</sup>[Honoré and de Paula \(2014\)](#) with SHARE data find  $\tau \approx 0.77$  and  $\delta \approx 1.0$ , using a different sample selection, controls, reference-age definition, and estimation method.

Variable (wives)	Coef.	Variable (husbands)	Coef.
$\alpha_w$	1.756 (0.098)	$\alpha_h$	1.603 (0.072)
Constant	-8.333 (0.404)	Constant	-6.492 (0.295)
Age <sub>t0</sub>	0.939 (0.085)	Age <sub>t0</sub>	0.691 (0.06)
Age gap	-0.049 (0.025)	Age gap	0.037 (0.016)
Primary edu	0.051 (0.184)	Primary edu	0.554 (0.221)
Tertiary edu	-0.163 (0.135)	Tertiary edu	-0.462 (0.109)
Poor health	-0.27 (0.139)	Poor health	-0.169 (0.126)
Number of grandchildren	0.096 (0.039)	Number of grandchildren	0.025 (0.024)
Number of children	-0.003 (0.051)	Number of children	-0.021 (0.044)
HH wealth	-0.681 (0.512)	HH wealth	-0.301 (0.525)
HH wealth squared	0.016 (1.158)	HH wealth squared	0.007 (1.218)
South Europe	-1.009 (0.254)	South Europe	-1.215 (0.187)
East Europe	0.586 (0.159)	East Europe	-0.765 (0.133)
North Europe	-0.7 (0.125)	North Europe	-1.249 (0.125)
$\delta$	1.062 (0.013)	$\tau$	0.732 (0.119)

**TABLE 1** Estimation results for the joint retirement model

2 Notes: The table reports parameter estimates with bootstrap standard errors (100 replications) in parentheses. The left column shows results for wives, and the right column for husbands. Both age at  $t = 0$  and HH wealth measures are standardized. The discount rate is set to  $\rho = 0.004$  and threat points are set at 60% of the maximum utility an individual would achieve without complementarity, following [Honoré and de Paula \(2018\)](#).

To further interpret the complementarity effect, we conduct a simple counterfactual experiment by comparing the estimated model with a scenario without complementarity ( $\delta = 1$ ). The results suggest that complementarity reduces the median retirement age by approximately 5 months for wives and 2 months for husbands.

## 5 | CONCLUSIONS

This paper studies optimization and estimation methods for choice models with piecewise smooth objective functions. Specifically, we establish conditions under which a choice model can be transformed into a reduced form resembling a censored model, whose potential outcomes are well-defined latent variables derived from auxiliary functions. This transformation not only provides an alternative solution method but also facilitates the implementation of the SEM algorithm for the estimation.

We present two examples to illustrate our methods: 1) responses to tax and transfers, and 2) the  $(S, s)$  model. Additionally, by applying our methods, we empirically estimate the interdependent duration model in Honoré and de Paula (2018) using the SHARE data to study joint retirement decisions among European couples. The results provide empirical evidence of complementarities between couples in their retirement processes.

## ACKNOWLEDGEMENTS

We thank Dante Amengual, Manuel Arellano, Dmitry Arkhangelsky, Stéphane Bonhomme, Angela Denis, Aureo de Paula, Micolé de Vera, Juan F. Jimeno, Pedro Gete, Laura Hospido, Elena Manresa, Pedro Mira, Liyang Sun and seminar participants at IE University, CEMFI, Pontificia Universidad Católica de Chile, UIBE, SAEe, Encounters in Econometric Theory, Barcelona Summer Forum, IAAE, and ESWC meetings for very helpful comments and suggestions. This research was generously supported by the research grant from XXIII Concurso Nacional para la Adjudicación de Ayudas a la Investigación en Ciencias Sociales, Fundación Ramón Areces and Agencia Nacional de Investigación y Desarrollo (FONDECYT-11220506).

## REFERENCES

- Arcidiacono, P. and Jones, J. B. (2003) Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, **71**, 933–946.
- Arellano, M. and Bonhomme, S. (2016) Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal*, **19**, C61–C94. URL: <http://www.jstor.org/stable/45172103>.
- Arrow, K. J., Harris, T. and Marschak, J. (1951) Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 250–272.
- Attanasio, O., Larkin, K., Ravn, M. O. and Padula, M. (2022)  $(s, s)$  cars and the great recession. *Econometrica*, **90**, 2319–2356.
- Attanasio, O. P. (2000) Consumer durables and inertial behaviour: Estimation and aggregation of  $(s, s)$  rules for automobile purchases. *The Review of Economic Studies*, **67**, 667–696.
- Banks, J., Blundell, R. and Rivas, M. C. (2007) The dynamics of retirement behavior in couples: Reduced-form evidence from england and the us. *University College London, mimeo* ([http://m.rand.org/content/dam/rand/www/external/labor/aging/rsi/rsi\\_papers/2008/banks1.pdf](http://m.rand.org/content/dam/rand/www/external/labor/aging/rsi/rsi_papers/2008/banks1.pdf)).

- Bertanha, M., McCallum, A. H. and Seegert, N. (2023) Better bunching, nicer notching. *Journal of Econometrics*, **237**, 105512.
- Blau, D. M. (1998) Labor force dynamics of older married couples. *Journal of Labor Economics*, **16**, 595–629.
- Blau, D. M. and Gilleskie, D. B. (2006) Health insurance and retirement of married couples. *Journal of Applied Econometrics*, **21**, 935–953.
- Blomquist, S., Hausman, J. A. and Newey, W. K. (2023) The econometrics of nonlinear budget sets. *Annual Review of Economics*, **15**, 287–306.
- Blomquist, S., Kumar, A., Liang, C.-Y. and Newey, W. K. (2015) Individual heterogeneity, nonlinear budget sets, and taxable income.
- Blomquist, S., Newey, W. K., Kumar, A. and Liang, C.-Y. (2021) On bunching and identification of the taxable income elasticity. *Journal of Political Economy*, **129**, 2320–2343.
- Blundell, R., French, E. and Tetlow, G. (2016) Retirement incentives and labor supply. In *Handbook of the economics of population aging*, vol. 1, 457–566. Elsevier.
- Bover, O. (2010) Housing purchases and the dynamics of housing wealth.
- Caballero, R. J. (1993) Durable goods: An explanation for their slow adjustment. *Journal of Political Economy*, **101**, 351–384.
- Caballero, R. J. and Engel, E. M. (1999) Explaining investment dynamics in us manufacturing: a generalized (s, s) approach. *Econometrica*, **67**, 783–826.
- Celeux, G., Chauveau, D. and Diebolt, J. (1996) Stochastic versions of the em algorithm: an experimental study in the mixture case. *Journal of statistical computation and simulation*, **55**, 287–314.
- Chernozhukov, V. and Hong, H. (2003) An mcmc approach to classical estimation. *Journal of Econometrics*, **115**, 293–346. URL: <https://www.sciencedirect.com/science/article/pii/S0304407603001003>.
- Chiappori, P.-A., Donni, O. and Komunjer, I. (2012) Learning from a piece of pie. *The Review of Economic Studies*, **79**, 162–195.
- Coile, C. (2003) Retirement incentives and couples' retirement decisions.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1–38. URL: <http://www.jstor.org/stable/2984875>.
- Eberly, J. C. (1994) Adjustment of consumers' durables stocks: Evidence from automobile purchases. *Journal of political Economy*, **102**, 403–436.
- Forneron, J.-J. (2023) Noisy, non-smooth, non-convex estimation of moment condition models. *arXiv preprint arXiv:2301.07196*.
- García-Miralles, E. and Leganza, J. M. (2024) Joint retirement of couples: Evidence from discontinuities in denmark. *Journal of Public Economics*, **230**, 105036.
- Gouriéroux, C. and Monfort, A. (1997) *Simulation-based Econometric Methods*. Oxford University Press. URL: <https://doi.org/10.1093/0198774753.001.0001>.
- Grossman, S. and Laroque, G. (1990) Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica*, **58**, 25–51.
- Gustman, A. L. and Steinmeier, T. L. (2000) Retirement in dual-career families: a structural model. *Journal of Labor economics*, **18**, 503–545.
- (2004) Social security, pensions and retirement behaviour within the family. *Journal of Applied Econometrics*, **19**, 723–737.

- Honoré, B. E. and de Paula, Á. (2014) *Joint retirement in europe*. SSRN.
- (2018) A new model for interdependent durations. *Quantitative Economics*, **9**, 1299–1333.
- Hospido, L. and Zamarro, G. (2014) Retirement patterns of couples in europe. *IZA Journal of European Labor Studies*, **3**, 1–18.
- Hurd, M. D. (1990) The joint retirement decision of husbands and wives. In *Issues in the Economics of Aging*, 231–258. University of Chicago Press, 1990.
- Johnsen, J. V., Vaage, K. and Willén, A. (2022) Interactions in public policies: Spousal responses and program spillovers of welfare reforms. *The Economic Journal*, **132**, 834–864.
- Johnsen, J. V. and Willén, A. (2022) The effect of negative income shocks on pensioners. *Labour Economics*, **76**, 102175.
- Karadi, P., Nakov, A., Nuño, G., Pasten, E. and Thaler, D. (2025) Strike while the iron is hot: optimal monetary policy with a nonlinear phillips curve.
- Van der Klaauw, W. and Wolpin, K. I. (2008) Social security and the retirement and savings behavior of low-income households. *Journal of econometrics*, **145**, 21–42.
- Lalive, R. and Parrotta, P. (2017) How does pension eligibility affect labor supply in couples? *Labour Economics*, **46**, 177–188.
- Michaud, P.-C., Van Soest, A. and Bissonnette, L. (2020) Understanding joint retirement. *Journal of Economic Behavior Organization*, **173**, 386–401. URL: <https://www.sciencedirect.com/science/article/pii/S0167268119302331>.
- Michaud, P.-C. and Vermeulen, F. (2011) A collective labor supply model with complementarities in leisure: Identification and estimation by means of panel data. *Labour Economics*, **18**, 159–167.
- Newey, W. K. and McFadden, D. (1994) Large sample estimation and hypothesis testing. *Handbook of econometrics*, **4**, 2111–2245.
- Nielsen, S. F. (2000) The stochastic em algorithm: estimation and asymptotic results. *Bernoulli*, 457–489.
- Pozzoli, D. and Ranzani, M. (2009) Old european couples' retirement decisions: The role of love and money.
- Pudney, S. (1989) *Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes*. Blackwell. URL: <https://books.google.es/books?id=zvpnQgAACAAJ>.
- Richard, M. (2024) Durable consumption during the great recession: the role of risk heterogeneity.
- Rosenblatt, M. (1952) Remarks on a multivariate transformation. *The annals of mathematical statistics*, **23**, 470–472.
- Saez, E. (2010) Do taxpayers bunch at kink points? *American economic Journal: economic policy*, **2**, 180–212.
- Scarf, H. (1960) The optimality of (s, s) policies in the dynamic inventory problem. *Mathematical Methods in the Social Sciences*.
- Smith Jr, A. A. (1993) Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, **8**, S63–S84.
- Wei, S. (2024) Estimating latent-variable panel data models using parameter-expanded sem methods. *Journal of Business & Economic Statistics*, 1–14.

## Appendix

### A | SIMULATION RESULTS

In this section, we present the simulation results for the 1) kinked budget constraint and 2) (S, s) model discussed in Section 3.

#### A.1 | Kinked Budget Constraint Simulation Results

We conduct two simulation exercises. In both exercises, the DGP is the choice model described in Section A. In the first exercise we allow for two kinks by setting three different local marginal tax rates,  $t_0 = 0.2$ ,  $t_1 = 0.3$ ,  $t_2 = 0.5$ , separated by thresholds  $\tau_1 = 1.5$  and  $\tau_2 = 2.5$ . We set the true elasticity parameter  $\theta = 0.3$ , and draw  $u$  from a log-normal distribution,  $\log(u) \sim N(0, \sigma^2)$ , where  $\sigma = 0.8$ . In the M-step, we can update parameter values by  $\hat{\theta}^{(s+1)} = \gamma(y_1^*) / \log(1 - t_1)$  and  $\hat{\sigma} = \gamma(y_1^*)$ .

Figure A1 presents the estimation results of one simulation with  $N = 5000$ . We plot the M-step updates of  $\hat{\theta}$  and  $\hat{\sigma}$  for 100 iterations. The orange solid line depicts each update of SEM starting from a set of random initial guess, whereas the blue dashed line represents the true value. It is straightforward that SEM updates moves immediately toward the true value and the final estimates are the average of the last 20 iterations, represented by the orange diamond. The whole process takes less than one second.

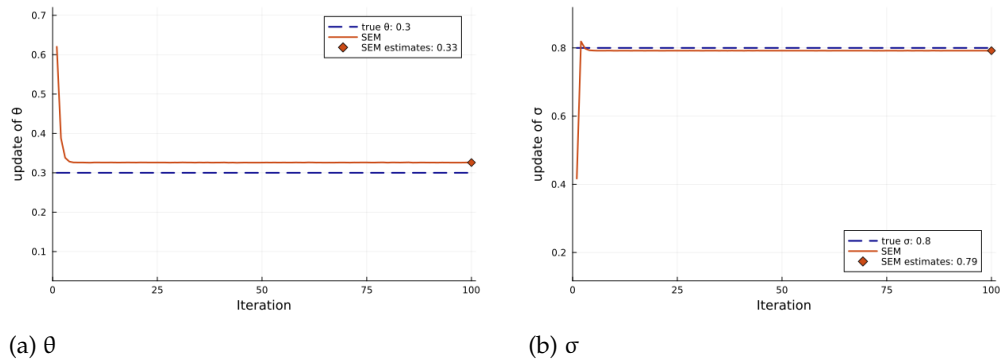
In the second exercise, we add covariates such that the utility functions takes the form

$$U(c, y; u, \theta) = c - \frac{u \exp(x'\beta)}{1 + 1/\theta} \left( \frac{y}{u \exp(x'\beta)} \right)^{1+1/\theta}$$

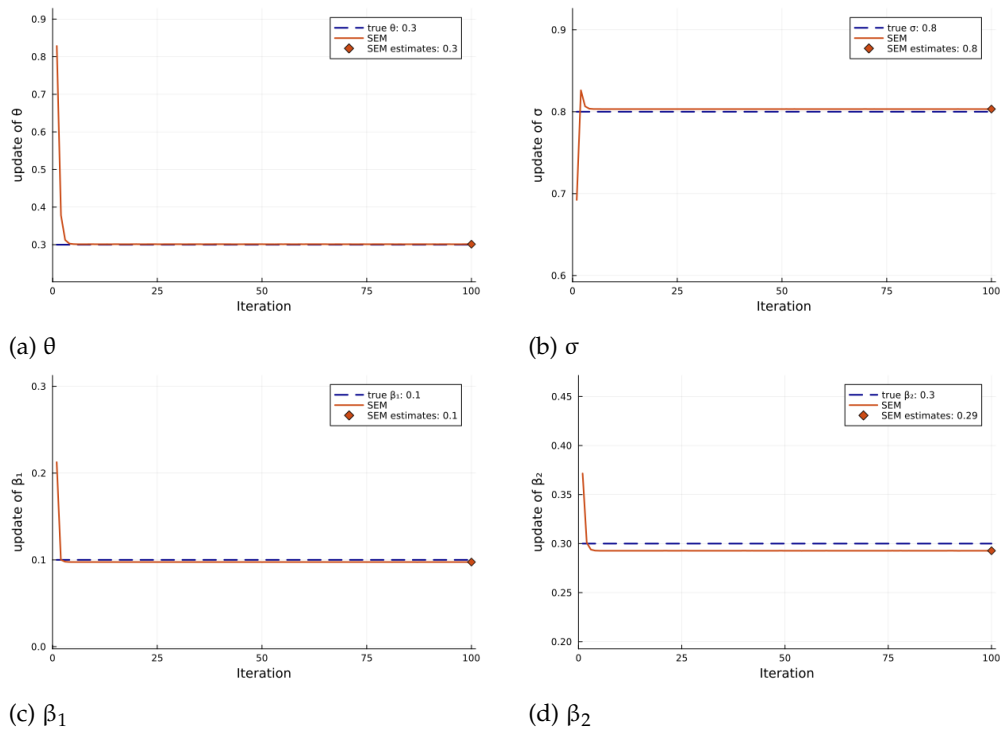
following Bertanha et al. (2023). We set  $t_0 = 0.2$ ,  $t_1 = 0.3$ ,  $\tau = 1.5$ ,  $\beta_1 = 0.1$ ,  $\beta_2 = 0.3$ , and true elasticity parameter  $\theta = 0.3$ . Similarly, draw  $u$  from a log-normal distribution,  $\log(u) \sim N(0, \sigma^2)$ , where  $\sigma = 0.8$ . In the M-step, we update the parameter values of  $\beta$  and  $\theta$  by regressing  $[y_1^* \ y_2^*]'$  on  $[\mathbf{1}_2 \otimes x \ \log(1 - t_0), \log(1 - t_1)]' \otimes \mathbf{1}_N$  and estimate  $\sigma$  as the standard deviation of the resulting residuals, where  $\mathbf{1}_m$  denotes an  $m \times 1$  vector of ones and  $\otimes$  denotes the Kronecker product.

Figure A2 presents the estimation results of one simulation with  $N = 5000$ . We observe a similar pattern as previous case, that is all M-step updates move immediately toward to the true value and converges after only a few iterations. In terms of total running time, the whole process takes less than one second.<sup>1</sup>

<sup>1</sup>The results are obtained using a Mac Studio (Apple M2 Ultra) with a single processor core.



**FIGURE A1** SEM iteration of  $\hat{\theta}$  and  $\hat{\sigma}$  from a random initial guess  
 2 Note: Iterations of SEM (orange line) based on direct sampling, compared with the true value (blue dashed line). SEM estimates (orange diamond) are calculated as the average of the last 20 iterations.  $N = 5000$ .



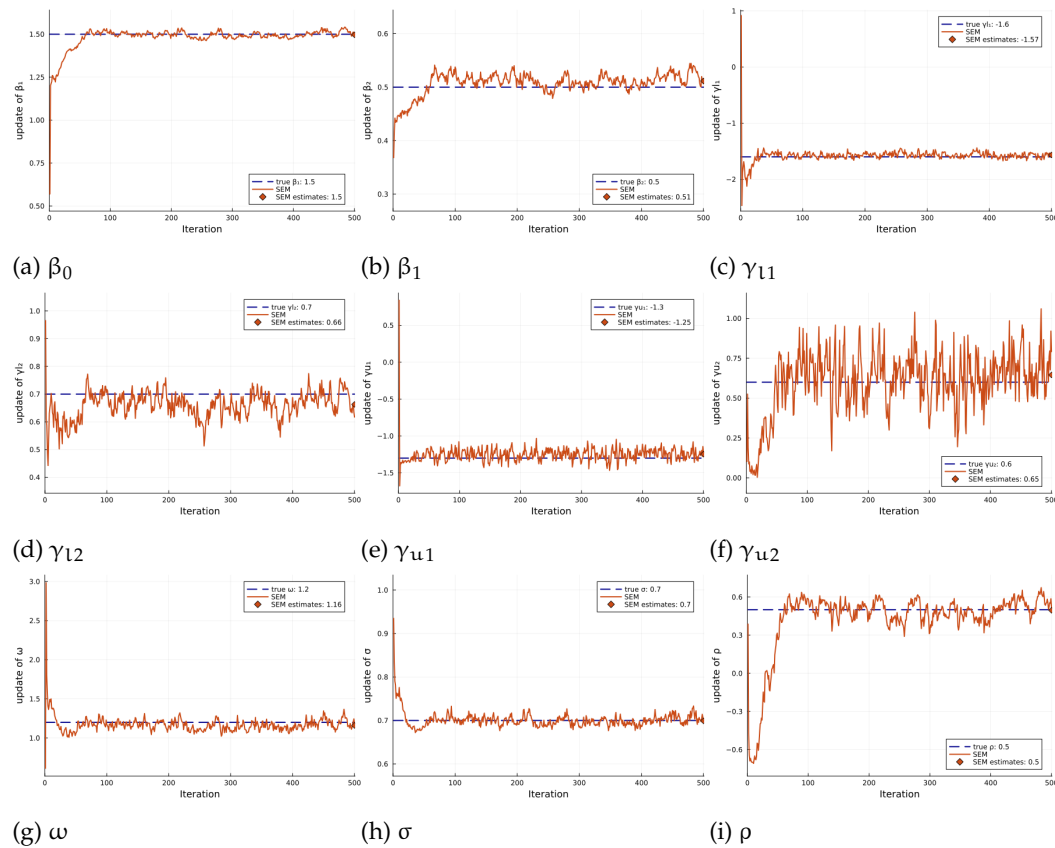
**FIGURE A2** SEM iteration of  $\theta$ ,  $\sigma$ , and  $\beta$  from a random initial guess  
 2 Note: Iterations of SEM (orange line) based on direct sampling, compared with the true value (blue dashed line). SEM estimates (orange diamond) are calculated as the average of the last 20 iterations.  $N = 5000$ .

## A.2 | (S, s) Model Simulation Results

We conduct a simulation exercise of the (S, s) model based on the specification in Section 3, using the following true parameter values:  $\beta = [1.5; 0.5]$ ,  $\gamma_\ell = [-1.6; 0.7]$ ,  $\gamma_u = [-1.3; 0.6]$ ,  $\omega = 1.2$ ,  $\sigma = 0.7$ , and  $\rho = 0.5$ , with a sample size of  $N = 5000$ . To complete the model, we assume that  $x = z \sim N(0, 1)$ .

Starting from a set of random initial guesses (i.e., draws from a uniform distribution), we perform the SEM algorithm for 500 iterations, following the procedure proposed in Section 3. Specifically, in the M-step, instead of maximizing the joint likelihood, we update parameters separately. Treating the E-step draws  $y^*$  as observables, we estimate  $\beta$  and  $\sigma$  by linear regression, and obtain the residuals  $\hat{u}$ . Next, we update  $\gamma_\ell$  and  $\gamma_u$  through two separate Probit regressions of  $\mathbb{1}(d \neq 0)$  on  $z$ ,  $\log(|y^* - y_1|)$ , and  $\hat{u}$ , using the subsamples with  $y^* < y_1$  and  $y^* > y_1$ , respectively. Finally, we update  $\rho$  and  $\omega$  based on a Probit regression using the full sample, given  $\gamma_\ell$  and  $\gamma_u$ , where the Probit regression is of  $\mathbb{1}(d \neq 0)$  on  $\log(|y^* - y_1|) - z\hat{\gamma}_\ell\mathbb{1}(y^* > y_1) - z\hat{\gamma}_u\mathbb{1}(y^* < y_1)$  and  $\hat{u}$ .

Figure A3 presents the estimation results of our simulation exercise. Similarly, the orange solid line depicts each update of SEM M-step updates, whereas the blue dashed line represents the true value. From the figure, we can see that all the updates approaches to the true values within 100 iterations. We take the average of the last 300 iterations as our final estimates. The whole process for running 500 iterations takes around 25 seconds.<sup>2</sup>



**FIGURE A3** SEM iteration of  $\theta$ ,  $\sigma$ , and  $\beta$  from a random initial guess  
<sup>2</sup> Note: Iterations of SEM (orange line) based on direct sampling, compared with the true value (blue dashed line). SEM estimates (orange diamond) are calculated as the average of the last 300 iterations.  $N = 5000$ .

<sup>2</sup>The results are obtained using a Mac Studio (Apple M2 Ultra) with a single processor core.

## B | AUXILIARY FUNCTIONS AND VARIABLES IN HDP MODEL

In this section, we verify that the following objective function of the HDP model satisfies Assumptions 1-3:

$$V(t_w, t_h) = \left( G_w(t_w) + (\delta - 1)\tilde{H}_w(\max(t_w, t_h)) - A_w \right) \\ \times \left( G_h(t_h) + (\delta - 1)\tilde{H}_h(\max(t_w, t_h)) - A_h \right).$$

where  $\tilde{H}_i(t) \equiv \int_t^\infty H_i(s)e^{-\rho s} ds$  and  $G_i(t) \equiv \int_0^t k_i e^{-\rho s} ds + \int_t^\infty H_i(s)e^{-\rho s} ds = k_i \rho^{-1}(1 - e^{-\rho t}) + \tilde{H}_i(t)$ .

Throughout this appendix we assume that, for each  $i \in \{w, h\}$ , the function  $H_i(t)$  is continuous, strictly increasing with  $H_i(0) = 0$ , and satisfies  $\lim_{t \rightarrow \infty} H_i(t) = \infty$ . Under these conditions, for any  $k_i > 0$  and  $\delta \geq 1$ , the functions  $k_i - H_i(t)$  and  $k_i - \delta H_i(t)$  each have a unique root on  $(0, \infty)$ .

### Feasible set.

We first define the feasible set  $S_y$ . Besides the basic conditions  $t_w > 0$  and  $t_h > 0$ , we must ensure that both components of  $V(t_w, t_h)$  are positive. Accordingly, in the region  $t_w \geq t_h > 0$ , we require  $G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w > 0$  and  $G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h > 0$ ; whereas in the region  $0 < t_w < t_h$ , we require  $G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h > 0$  and  $G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w > 0$ . Hence,

$$S_y \equiv \left\{ (t_w, t_h) : t_w \geq t_h > 0, G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w > 0, \right. \\ \left. G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h > 0 \right\} \\ \cup \left\{ (t_w, t_h) : 0 < t_w < t_h, G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h > 0, \right. \\ \left. G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w > 0 \right\}.$$

Moreover, since  $\tilde{H}_i(t)$  is monotonically decreasing in  $t$ , we have, in the region  $t_w \geq t_h > 0$ ,

$$G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w \geq G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w > 0, \\ G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h \geq G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h > 0.$$

Similarly, in the region  $0 < t_w < t_h$ , we have

$$G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h \geq G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h > 0, \\ G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w \geq G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w > 0.$$

Therefore, within the feasible set  $S_y$ , we have

$$G_i(t_i) + (\delta - 1)\tilde{H}_i(t_i) - A_i > 0 \quad \text{and} \quad G_i(t_i) + (\delta - 1)\tilde{H}_i(t_{-i}) - A_i > 0, \quad \forall i \in \{w, h\}.$$

### Assumption 1.

This assumption imposes a regularity condition ensuring the existence of a unique maximizer within the feasible set  $S_y$ . Since  $V(t_w, t_h)$  is continuous on  $S_y$  and has a unique maximizer in each region (as implied by Assumptions 2-3 in the remaining part), and since

the objective value depends on the random shocks  $k_w$  and  $k_h$ , the global maximizer exists and is unique almost surely.

**Assumption 2.**

We now turn to Assumption 2. Specifically, we verify that the following two auxiliary functions are unimodal and characterize their global optimizers:

$$\begin{aligned} V_w(t_w, t_h) &\equiv (G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w)(G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h), \\ V_h(t_w, t_h) &\equiv (G_w(t_w) + (\delta - 1)\tilde{H}_w(t_h) - A_w)(G_h(t_h) + (\delta - 1)\tilde{H}_h(t_h) - A_h). \end{aligned}$$

We take  $V_w(t_w, t_h)$  as an example; the same reasoning applies to  $V_h(t_w, t_h)$ . For the auxiliary function  $V_w(t_w, t_h)$ , the derivative with respect to  $t_h$  is

$$\frac{\partial V_w(t_w, t_h)}{\partial t_h} = (k_h e^{-\rho t_h} - H_h(t_h) e^{-\rho t_h}) \times (G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w).$$

One can verify that, within the subset  $S_y$  where the second term remains positive, and for any given  $t_w$ , as  $t_h$  increases from its smallest feasible value in  $S_y$ , the function  $V_w(t_w, t_h)$  first increases and then decreases monotonically along the  $t_h$  direction.

The derivative with respect to  $t_w$  is slightly more involved:

$$\begin{aligned} \frac{\partial V_w(t_w, t_h)}{\partial t_w} &= \underbrace{(k_w e^{-\rho t_w} - \delta H_w(t_w) e^{-\rho t_w})}_{\text{(i) as } t_w \text{ increases, } >0 \rightarrow =0 \rightarrow <0} \times \underbrace{(G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h)}_{\text{(ii)} >0} \\ &+ \underbrace{(G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w)}_{\text{(iii)} >0} \times \underbrace{(-(\delta - 1)H_h(t_w) e^{-\rho t_w})}_{\text{(iv)} \leq 0}. \end{aligned}$$

Within the subset  $S_y$ , for any given value of  $t_h$ , the term (i)  $\times$  (ii) changes from positive to negative as  $t_w$  increases, whereas the term (iii)  $\times$  (iv) remains negative. Moreover, when  $t_w$  is small,  $H_h(t_w)$  is close to zero, while  $k_w - \delta H_w(t_w)$  is close to  $k_w$ , and thus term (i) dominates term (iv). Therefore, for a fixed  $t_h$ , as  $t_w$  increases from its smallest feasible value, the function  $V_w(t_w, t_h)$  first increases and then decreases.

We now characterize its optimizer over the domain  $S_y$ . The optimizer  $op_w \equiv (T_w^w, T_h^w)$  needs to satisfy

$$\frac{\partial V_w(T_w^w, T_h^w)}{\partial t_w} = 0 \quad \text{and} \quad \frac{\partial V_w(T_w^w, T_h^w)}{\partial t_h} = 0.$$

Moreover, since  $G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w > 0$  on  $S_y$ , the first-order condition with respect to  $t_h$  simplifies to  $k_h = H_h(T_h^w)$ , and the first-order condition can be written compactly as

$$g_w(op_w) \equiv \begin{bmatrix} T_h^w - H_h^{-1}(k_h) \\ \psi(T_w^w, T_h^w, x_w, x_h, k_w, k_h) \end{bmatrix} = \mathbf{0},$$

where  $\psi(t_w, t_h, x_w, x_h, k_w, k_h) \equiv H_h(t_w) e^{-\rho t_w} (1 - \delta) \times (G_w(t_w) + (\delta - 1)\tilde{H}_w(t_w) - A_w) + (G_h(t_h) + (\delta - 1)\tilde{H}_h(t_w) - A_h) \times (k_w e^{-\rho t_w} - H_w(t_w) \delta e^{-\rho t_w})$ .

**Assumption 3.**

The objective function on the common boundary is

$$V_j(t, t) \equiv (G_w(t) + (\delta - 1)\tilde{H}_w(t) - A_w)(G_h(t) + (\delta - 1)\tilde{H}_h(t) - A_h).$$

It follows that the first-order condition is:

$$\begin{aligned} \frac{\partial V_j(t, t)}{\partial t} &\equiv (k_w e^{-\rho t} - \delta H_w(t) e^{-\rho t})(G_h(t) + (\delta - 1)\tilde{H}_h(t) - A_h) \\ &\quad + (G_w(t) + (\delta - 1)\tilde{H}_w(t) - A_w)(k_h e^{-\rho t} - \delta H_h(t) e^{-\rho t}). \end{aligned}$$

Starting from small values of  $t$ , the function  $V_j(t, t)$  first increases and then decreases. The optimizer  $op_j \equiv (T^j, T^j)$  for the function  $V_j(t, t)$  must satisfy

$$g_j(op_j) \equiv \eta(T^j, T^j, x_w, x_h, k_w, k_h) = 0,$$

where  $\eta(t, t, x_w, x_h, k_w, k_h) = (k_w - \delta H_w(t))(G_h(t) + (\delta - 1)\tilde{H}_h(t) - A_h) + (k_h - \delta H_h(t))(G_w(t) + (\delta - 1)\tilde{H}_w(t) - A_w)$ .

## C | SEM PROCEDURES AND SIMULATIONS FOR HDP MODEL

In this appendix, we begin with explaining the SEM algorithm for the HDP model and then present simulation results.

### C.1 | SEM algorithm

First, we detail the SEM procedure to draw from the following framework:

$$g_w(op_w) \equiv \left[ \begin{array}{c} T_h^w - H_h^{-1}(k_h; \alpha_h, \beta_h) \\ \psi(T_w^w, T_h^w, x_w, x_h, k_w, k_h; \alpha_w, \beta_w, \alpha_h, \beta_h, \delta) \end{array} \right] = \mathbf{0}, \quad (C1)$$

$$g_h(op_h) \equiv \left[ \begin{array}{c} T_w^h - H_w^{-1}(k_w; \alpha_w, \beta_w) \\ \psi(T_h^h, T_w^h, x_h, x_w, k_h, k_w; \alpha_h, \beta_h, \alpha_w, \beta_w, \delta) \end{array} \right] = \mathbf{0}, \quad (C2)$$

$$g_j(op_j) \equiv \eta(T^j, T^j, x_w, x_h, k_w, k_h; \alpha_w, \beta_w, \alpha_h, \beta_h, \delta) = 0, \quad (C3)$$

$$(T_w, T_h) = op_w \mathbb{1}(T_w > T_h) + op_h \mathbb{1}(T_w < T_h) + op_j \mathbb{1}(T_w = T_h), \quad (C4)$$

where  $(k_w, k_h)$  follow  $\text{Exp}(1)$  marginal distributions with Clayton copula dependence parameter  $\tau$ .

The unknown parameter vector is  $\kappa \equiv [\alpha_w, \beta_w', \alpha_h, \beta_h', \delta, \tau]'$ , and the latent variables are  $op_w = (T_w^w, T_h^w)$ ,  $op_h = (T_w^h, T_h^h)$ , and  $op_j = (T^j, T^j)$ . In the E-step, given the current parameter guess  $\hat{\kappa}^{(s)}$ , we draw  $op_w$ ,  $op_h$ , and  $op_j$  from their posterior  $f(op_w, op_h, op_j | T_w, T_h, x_w, x_h; \hat{\kappa}^{(s)})$ . In the M-step, these draws, together with the observables, are used to update the parameters to  $\hat{\kappa}^{(s+1)}$ . Starting from  $\hat{\kappa}^{(0)}$ , we iterate the E- and M-steps for  $S$  iterations until  $\hat{\kappa}^{(s)}$  converges to a stationary distribution. The final estimator averages the last  $S_0$  iterates:  $\hat{\kappa} = \sum_{s=S-S_0+1}^S \hat{\kappa}^{(s)} / S_0$ .

#### E-step.

The sampling approach for  $op_w$ ,  $op_h$ , and  $op_j$  from their posterior distribution, given  $\hat{\kappa}^{(s)}$ , depends on the retirement type and on whether censoring is present.

*Sequential retirement, no censoring* ( $T_w > T_h$  or  $T_w < T_h$ ): Consider the case  $T_w > T_h$ . Equation (C4) immediately implies  $op_w = (T_w, T_h)$ . Given the current parameter guess  $\hat{\kappa}^{(s)}$ , and using  $g_w(op_w) = \mathbf{0}$ , we can solve for the unique pair  $(k_w, k_h)$ . Once  $k_w$  and  $k_h$  are obtained,  $op_h$  and  $op_j$  can be computed through  $g_h(op_h) = \mathbf{0}$  and  $g_j(op_j) = 0$ . For the opposite case,  $T_w < T_h$ , we have  $op_h = (T_w, T_h)$ . Using  $g_h(op_h) = 0$ , we similarly solve for the unique  $(k_w, k_h)$  and then compute  $op_w$  and  $op_j$ .

*Joint retirement, no censoring* ( $T_w = T_h$ ): Equation (C4) implies  $op_j = (T_w, T_h)$ , and thus  $k_w$  and  $k_h$  must satisfy  $g_j(op_j) = 0$ , which admits infinitely many combinations of  $(k_w, k_h)$ . However, natural lower and upper bounds exist for this set. Let the observed retirement time be  $T_w = T_h = t^*$ . The first boundary case occurs when the wife retires at  $T_w = t^*$ , and the husband retires almost simultaneously at  $T_h = t^* + \Delta$  with  $\Delta \rightarrow +0$ . The lower bound of  $k_w$ , denoted  $k_w^l$ , is obtained from  $g_h(t^*, t^*) = \mathbf{0}$ , specifically,

$$t^* = H_w^{-1}(k_w^l).$$

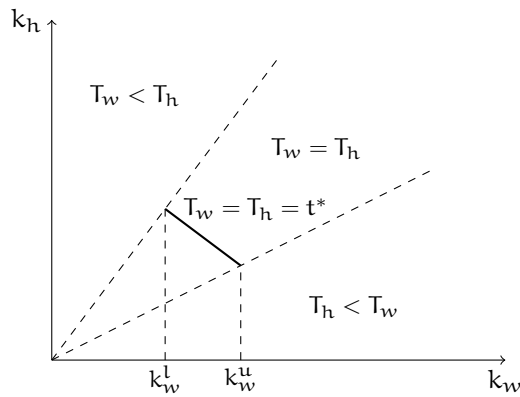
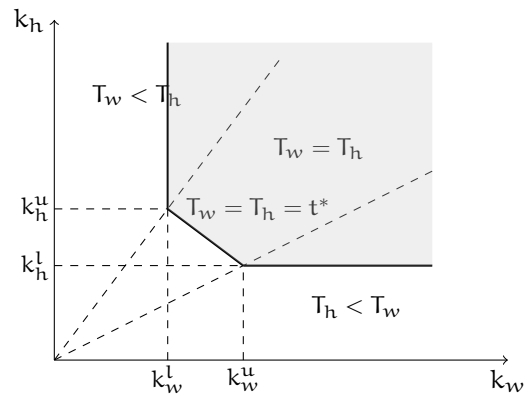
The other boundary case occurs when the husband retires first at  $T_h = t^*$ , and the wife retires almost simultaneously at  $T_w = t^* + \Delta$  with  $\Delta \rightarrow +0$ . The upper bound  $k_w^u$  is solved from  $g_w(t^*, t^*) = \mathbf{0}$ , specifically,

$$\psi(t^*, t^*, x_w, x_h, k_w^u, H_h(t^*)) = 0.$$

The left panel of Figure C1 illustrates these boundaries.

Then, for each household, we randomly draw one pair  $(k_w, k_h)$  that satisfies  $g_j(op_j) = 0$  with  $k_w^l \leq k_w \leq k_w^u$ , and with probabilities proportional to the joint distribution of  $(k_w, k_h)$ .

The random draw is implemented using the Metropolis–Hastings algorithm, discarding the first 100 draws as burn-in and targeting an acceptance rate between 20 – 40%. Once  $k_w$  and  $k_h$  are obtained,  $op_w$  and  $op_h$  are computed from the equations  $g_w(op_w) = \mathbf{0}$  and  $g_h(op_h) = \mathbf{0}$ .

**FIGURE C1** Feasible sets for  $(k_w, k_h)$ (a) Joint retirement,  $T_w = T_h = t^*$ (b) Right censored with  $T_w > t^*, T_h > t^*$ 

2 Notes: The figures illustrate retirement decisions based on  $(k_w, k_h)$  values. The blue line in the left figure (could be curved) indicates all possible  $(k_w, k_h)$  consistent with observables. The right figure depicts the region of  $(k_w, k_h)$  when both partners' retirement times are censored at  $t^*$ .

*Right censored:* When only one member's retirement time is censored (typically when only the first retirement is observed), we can immediately identify the unique  $k_i$  value for the early retiree and a lower bound for  $k_{-i}$  for the later retiree. For example, if we observe  $T_w = t_1^*$  and  $T_h > t_2^* \geq t_1^* = T_w$ , this data pattern implies that the latent optimizer must be  $op_h$ , which is characterized by  $g_h(op_h) = \mathbf{0}$ . We then solve for the unique  $k_w$  and the lower bound of  $k_h$ :

$$\begin{aligned} t_1^* &= H_w^{-1}(k_w) \\ \psi(t_2^*, t_1^*, x_h, x_w, k_h^l, k_w) &= 0. \end{aligned} \quad (C5)$$

Then, by direct sampling, we draw  $k_h$  from the truncated conditional distribution given  $k_w$ , that is  $f(k_h|k_w, k_h > k_h^l)$ . With  $(k_w, k_h)$  determined, we compute the implied  $op_w$  and  $op_h$  from  $g_w(op_w) = \mathbf{0}$  and  $g_h(op_h) = \mathbf{0}$ . The same strategy applies when the husband retires first and the wife's time is censored.

When both retirement times are censored ( $T_w > t^*, T_h > t^*$ ), the case may correspond to either sequential or joint retirement ( $T_w > T_h > t^*, T_h > T_w > t^*$ , or  $T_w = T_h > t^*$ ), as shown in the right panel of Figure C1. The boundaries correspond to three limiting cases:  $T_h > T_w = t^*$ ,  $T_w > T_h = t^*$ , and  $T_w = T_h = t^*$ . Solving for these boundaries yields the feasible region for  $(k_w, k_h)$ :

$$\{(k_w, k_h) \in \mathbb{R}_+^2 | k_w > k_w^l, k_h > r(k_w)\},$$

where

$$r(k_w) = \begin{cases} \bar{k}(k_w), & k_w^l < k_w < k_w^u, \\ k_h^l, & k_w \geq k_w^u, \end{cases}$$

and the components are pinned down by  $k_h^l = H_h(t^*)$ ,  $\psi(t^*, t^*, x_w, x_h, k_w^u, H_h(t^*)) = 0$ , and  $\eta(t^*, t^*, x_w, x_h, k_w, \bar{k}(k_w)) = 0$ .

To generate a random draw within this region, we use rejection sampling: draw  $(k_w, k_h)$  from the joint distribution on  $\mathbb{R}_+^2$  and accept the first draw that satisfies the region conditions. Once  $(k_w, k_h)$  are obtained, the implied  $op_w$  and  $op_h$  are computed from  $g_w(op_w) = \mathbf{0}$  and  $g_h(op_h) = \mathbf{0}$ .

#### M-step.

We now describe the procedure for obtaining  $\hat{\kappa}^{(s+1)}$  using both the data and the E-step draws.

*Update  $\hat{\alpha}_i^{(s+1)}$  and  $\hat{\beta}_i^{(s+1)}$ :* Using the relations  $T_h^w = H_h^{-1}(k_h)$  and  $T_w^h = H_w^{-1}(k_w)$  satisfied by  $op_w$  and  $op_h$ , we estimate the following log-linear specification by OLS:

$$\log(T_i^{-i}) = -\frac{\gamma}{\alpha_i} - \frac{1}{\alpha_i} x_i' \beta_i + \frac{1}{\alpha_i} (\log(k_i) + \gamma), \quad i \in \{w, h\},$$

where  $k_i$ , given its Exp(1) marginal distribution, satisfy  $(\log(k_i) + \gamma | x_i) = 0$ , and  $(\log(k_i) | x_i) = \pi^2/6$ , with  $\gamma$  denoting the Euler-Mascheroni constant.

Let  $\hat{\sigma}_{\varepsilon, i}$  denote the standard deviation of the OLS residuals, then we can estimate  $\hat{\alpha}_i^{(s+1)} = \pi / (\hat{\sigma}_{\varepsilon, i} \sqrt{6})$ . The coefficients  $\hat{\beta}_i^{(s+1)}$  are obtained by rescaling the OLS estimates by  $-\hat{\alpha}_i^{(s+1)}$  and adjusting the intercept by  $-\gamma$ .

*Update of  $\hat{\delta}^{(s+1)}$  and  $\hat{\tau}^{(s+1)}$ :* The copula-dependence parameter  $\tau$  and the complemen-

tarity parameter  $\delta$  are updated by matching the moment conditions implied by the joint distribution of  $(k_w, k_h)$ :

$$\hat{\delta}^{(s+1)} = \arg \min_{\delta} \text{obj} \left( \tilde{k}_w(\hat{\tau}(\delta)), \tilde{k}_h(\hat{\tau}(\delta)) \right).$$

The following subsections describe in turn the definitions of  $k_i(\delta)$ ,  $\hat{\tau}(\delta)$ ,  $\tilde{k}_i(\tau(\delta))$ , and the objective function  $\text{obj}(\cdot)$ .

- \*  $k_i(\delta)$ : Given  $\hat{\alpha}_i^{(s+1)}$  and  $\hat{\beta}_i^{(s+1)}$ , for each candidate value of  $\delta$  we recompute the residual pair  $(k_w, k_h)$  implied by the E-step draws. In the sequential-retirement case without censoring, the computation is straightforward. In the joint-retirement and censored cases, however,  $\delta$  enters  $g_w(\cdot)$ ,  $g_h(\cdot)$ , and  $g_j(\cdot)$ , so the feasible support of  $(k_w, k_h)$  changes with  $\delta$ . To preserve the validity of SEM, we therefore normalize the residuals so that their relative positions within the feasible region remain unchanged as  $\delta$  varies.
  - Sequential retirement no censoring: When  $T_w > T_h$ , the residuals  $(k_w, k_h)$  are obtained by solving  $g_w(\cdot) = 0$ ; When  $T_w < T_h$ , they are obtained by solving  $g_h(\cdot) = 0$ .
  - Joint retirement, no censoring: When  $\delta$  changes, the bounds  $[k_w^l, k_w^u]$  defining the feasible segment of  $k_w$ , and the corresponding range of  $k_h$ , also change. We therefore compute  $k_i(\delta)$  while preserving the relative position of the E-step draw within its original bounds:

$$k_w(\delta) = k_w^l(\delta) + (k_w^u(\delta) - k_w^l(\delta)) * q_w, \quad \eta(t^*, t^*, x_w, x_h, k_w(\delta), k_h(\delta)) = 0,$$

where

$$q_w = \frac{k_w - k_w^l(\hat{\delta}^{(s)})}{k_w^u(\hat{\delta}^{(s)}) - k_w^l(\hat{\delta}^{(s)})}$$

denotes the relative location of the E-step draw  $k_w$  within its original support  $[k_w^l(\hat{\delta}^{(s)}), k_w^u(\hat{\delta}^{(s)})]$ .

- Sequential retirement with censoring: Similarly, when  $\delta$  changes, the feasible support of  $k_i$  also shifts. For example, if  $T_w = t_1^*$  and  $T_h > t_2^* \geq t_1^* = T_w$ , the lower bound of  $k_h$ ,  $k_h^l(\delta)$ , obtained from Equation (C5), varies with  $\delta$ . We therefore preserve the relative position of the E-step draw with respect to the bound  $k_h^l(\hat{\delta}^{(s)})$ :

$$k_h(\delta) = k_h^l(\delta) + (k_h - k_h^l(\hat{\delta}^{(s)})).$$

- Joint censored: When both retirement times are censored, we normalize the residuals following the same logic, depending on which option is optimal in the E-step. If  $op_j$  is the optimal choice in the E-step, we then normalize following the joint-retirement case; if  $op_w$  or  $op_h$  is optimal, we apply the sequential-retirement normalization.

- \*  $\hat{\tau}(\delta)$ : Given  $k_w(\delta)$  and  $k_h(\delta)$ , we compute the implied Clayton-copula dependence parameter as

$$\hat{\tau}(\delta) = \frac{2\kappa_{\text{endall}}(k_w(\delta), k_h(\delta))}{1 - \kappa_{\text{endall}}(k_w(\delta), k_h(\delta))},$$

where  $\kappa_{\text{endall}}$  denotes Kendall's rank correlation.

- \*  $\tilde{k}_i(\hat{\tau}(\delta))$ : The joint moments of  $(k_w, k_h)$  under a nonzero Clayton copula parameter  $\tau$  lack closed-form expressions. Therefore, we apply the Rosenblatt transformation ([Rosenblatt](#),

1952) to map any dependent pair  $(k_w, k_h)$  into independent variables  $(\tilde{k}_w(\hat{\tau}), \tilde{k}_h(\hat{\tau}))$ . The transformed variables  $(\tilde{k}_w, \tilde{k}_h)$  are each marginally  $\text{Exp}(1)$  and mutually independent, whose joint moments are known in closed form.

- \*  $\text{obj}(\cdot)$ : We rely on the third conditional moments of  $\tilde{k}_w - \tilde{k}_h$  and  $\tilde{k}_w + \tilde{k}_h$ , and on their independence from  $x_w$  and  $x_h$ . Specifically, the objective function includes (i) the absolute value of the regression coefficients from regressing  $(\tilde{k}_w - \tilde{k}_h)^3$  and  $(\tilde{k}_w + \tilde{k}_h)^3$  on  $x$ ; and (ii) the absolute deviations of skewness  $(\tilde{k}_w - \tilde{k}_h)$  and skewness  $(\tilde{k}_w + \tilde{k}_h)$  from their theoretical values 0 and  $\sqrt{2}$ . Absolute values are used for numerical stability.

Once the value of  $\delta$  that minimizes the objective function is chosen, we obtain  $\hat{\delta}^{(s+1)}$  and the corresponding  $\hat{\tau}^{(s+1)} = \hat{\tau}(\hat{\delta}^{(s+1)})$  as the M-step updates.

*Extra step:* In practice, however, when the sample is small or right-censoring is extensive, the M-step updates for  $\delta$  often exhibit high volatility across iterations, and the resulting SEM estimates display greater dispersion around the true parameter values in finite samples. To improve stability and precision in such cases, we add an additional M-step update after the steps above, using an unconditional moment condition that requires the predicted share of joint retirements in the M-step to match the empirical proportion from the E-step draws. This additional condition both sharpens the identification of  $\delta$  and regularizes the M-step update.

Specifically, given  $\hat{\alpha}_i^{(s+1)}$ ,  $\hat{\beta}_i^{(s+1)}$ , and  $\hat{\tau}^{(s+1)}$ , for any candidate value of  $\delta$  we can compute, for each household, the probability of joint retirement by evaluating numerically the probability that  $(k_w, k_h)$  falls between the two dashed lines in Figure C1. We then choose the value of  $\delta$  such that the unconditional probability of joint retirement matches the empirical proportion of joint retirement in the E-step draws. After updating  $\hat{\delta}^{(s+1)}$ , we set  $\hat{\tau}^{(s+1)} = \hat{\tau}(\hat{\delta}^{(s+1)})$ . In practice, this additional step substantially reduces instability in the M-step and improves the precision of the SEM estimator without altering the asymptotic target. Although the additional step adds to the computation time relative to the baseline SEM, the overall algorithm remains highly efficient with multithreaded implementation. The following subsection presents results on the computing time.

These steps jointly define the SEM estimation algorithm used in the empirical analysis.

## C.2 | Simulation Exercises

This subsection presents Monte Carlo simulations comparing the baseline SEM and the augmented SEM under different sample sizes,  $N = 10,000$  and  $N = 1,000$ , as well as results on the computing time of the augmented SEM method.

Based on the framework in equations (C1)–(C4), the true parameter values are set to  $\delta = 1.5$ ,  $\tau = 0.5$ ,  $\alpha_w = 1.24$ ,  $\alpha_h = 1.25$ ,  $\beta_w = [-5, -1.4, -0.1]'$ , and  $\beta_h = [-4.7, 1.3, 0.2]'$ . We introduce random right censoring as follows: for each household, there is an 80% probability of receiving a censoring time drawn uniformly from  $\{12, 24, \dots, 240\}$ . Each individual is censored if their true retirement time exceeds this censoring time.

For each sample-size configuration, I generate 40 independent datasets. While for the large-sample censored case, we estimate the model using the standard SEM algorithm, for the small-sample case, we estimate using both the standard and the augmented SEM algorithms. Each SEM algorithm is iterated for 100 steps, and the final estimate is obtained by averaging the last 50 iterations. The distributions of the parameter estimates across the 40 simulations are plotted to assess finite-sample variation under different scenarios.

**Sample size**  $N = 10,000$ .

Figure C2 reports the histograms of the baseline SEM estimates across 40 simulations. The estimates for all parameters are centered around the true values (indicated by the dark green

vertical line), confirming the accuracy and stability of the baseline SEM in large samples in the presence of censoring.

**Sample size  $N = 1,000$ .**

Figure C3 presents the histograms of both the baseline and augmented SEM estimates across 40 simulations for  $N = 1,000$  with censoring. Compared to the large-sample case, the finite-sample variability of the baseline SEM increases notably for  $\delta$  and  $\tau$  when sample size decreases. In contrast, the augmented SEM, which incorporates an additional unconditional moment restriction on  $\delta$ , substantially reduces the dispersion of these estimates, improving small-sample precision.

**SEM iterations and computing time.**

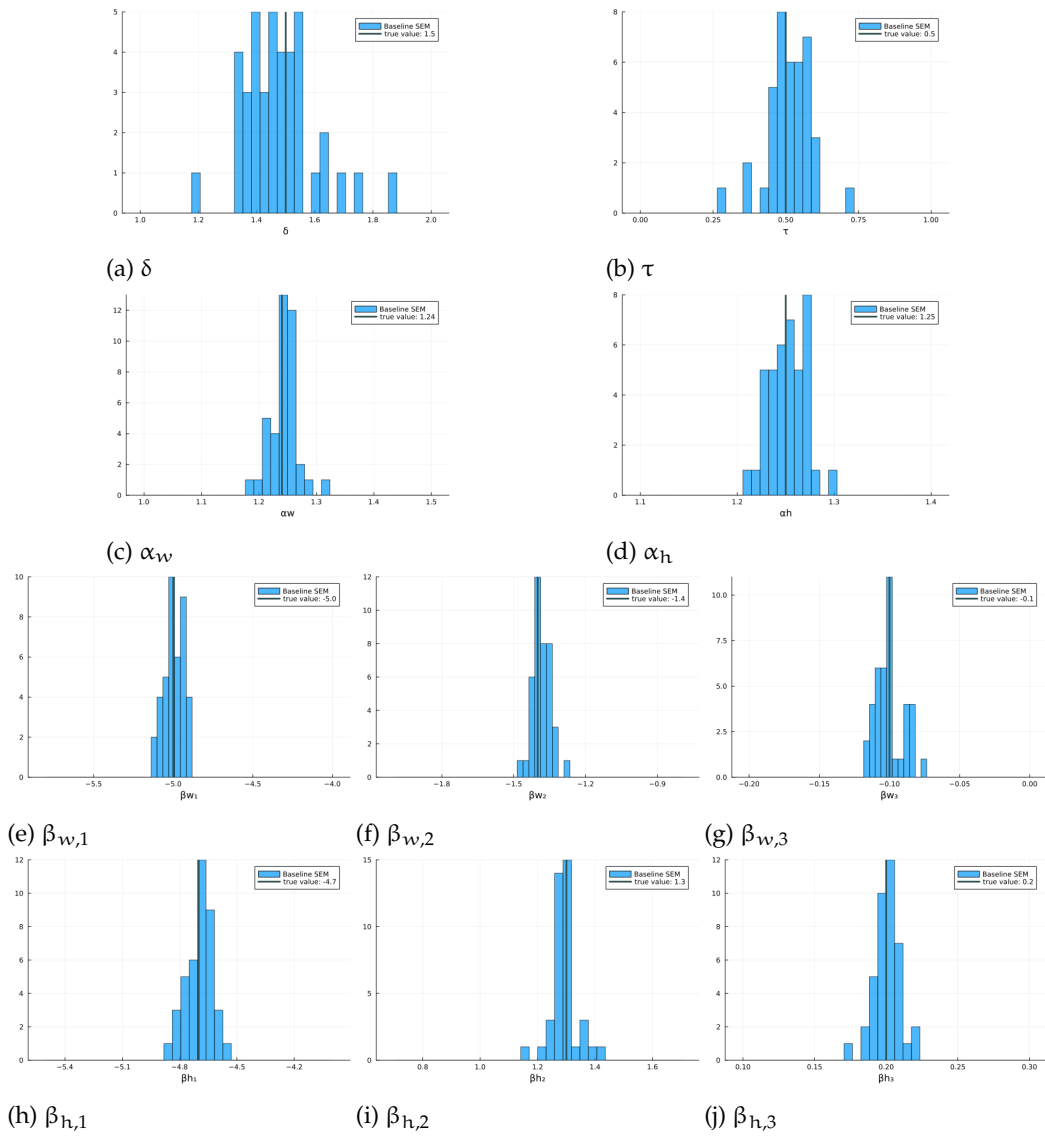
Finally, we present results on the computing time of the augmented SEM algorithm.

Figure C4 presents the M-step updates across iteration by computation time for the augmented SEM based on one simulated sample of size  $N = 1,000$ . The iteration starts from random initial values and runs for 100 steps. Each M-step update (solid blue line) approaches the true value (green dashed line) within approximately 30 seconds, and 100 iterations complete in about 4 minutes.<sup>3</sup>

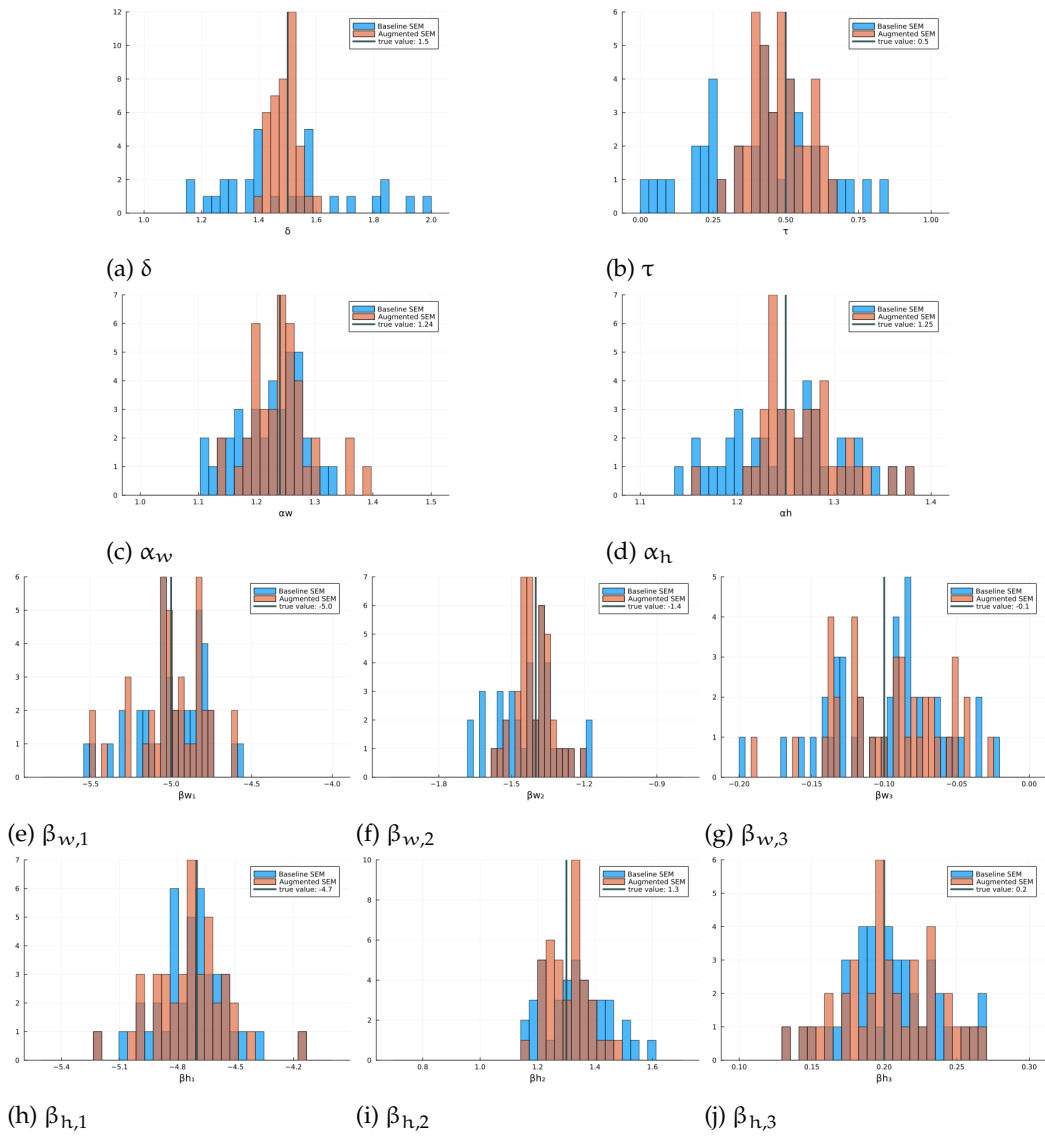
Similarly, Figure C5 presents the augmented SEM M-step updates over iterations by computation time for a larger sample size of  $N = 10,000$ . The results are based on 40 iterations starting from random initial guesses. As the sample size increases, the time required to complete each iteration also increases; consequently, completing 40 iterations takes around 16 minutes. However, as before, the M-step updates approaches to the true values very quickly—within just a few iterations — requiring only about 2 – 3 minutes.

---

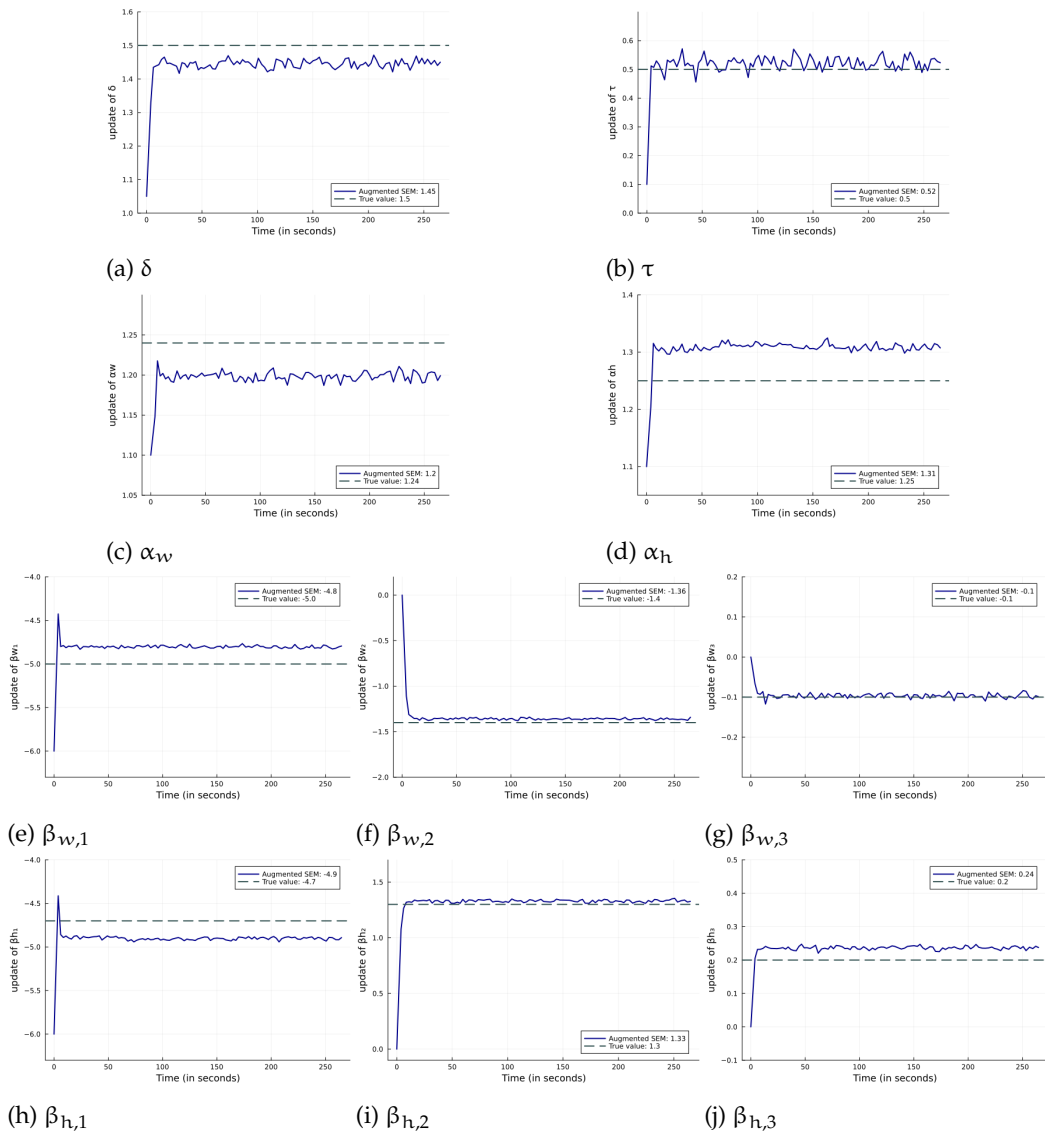
<sup>3</sup>Simulation run on Mac Studio M3 Ultra, multithreaded setup with 20 cores in Julia.



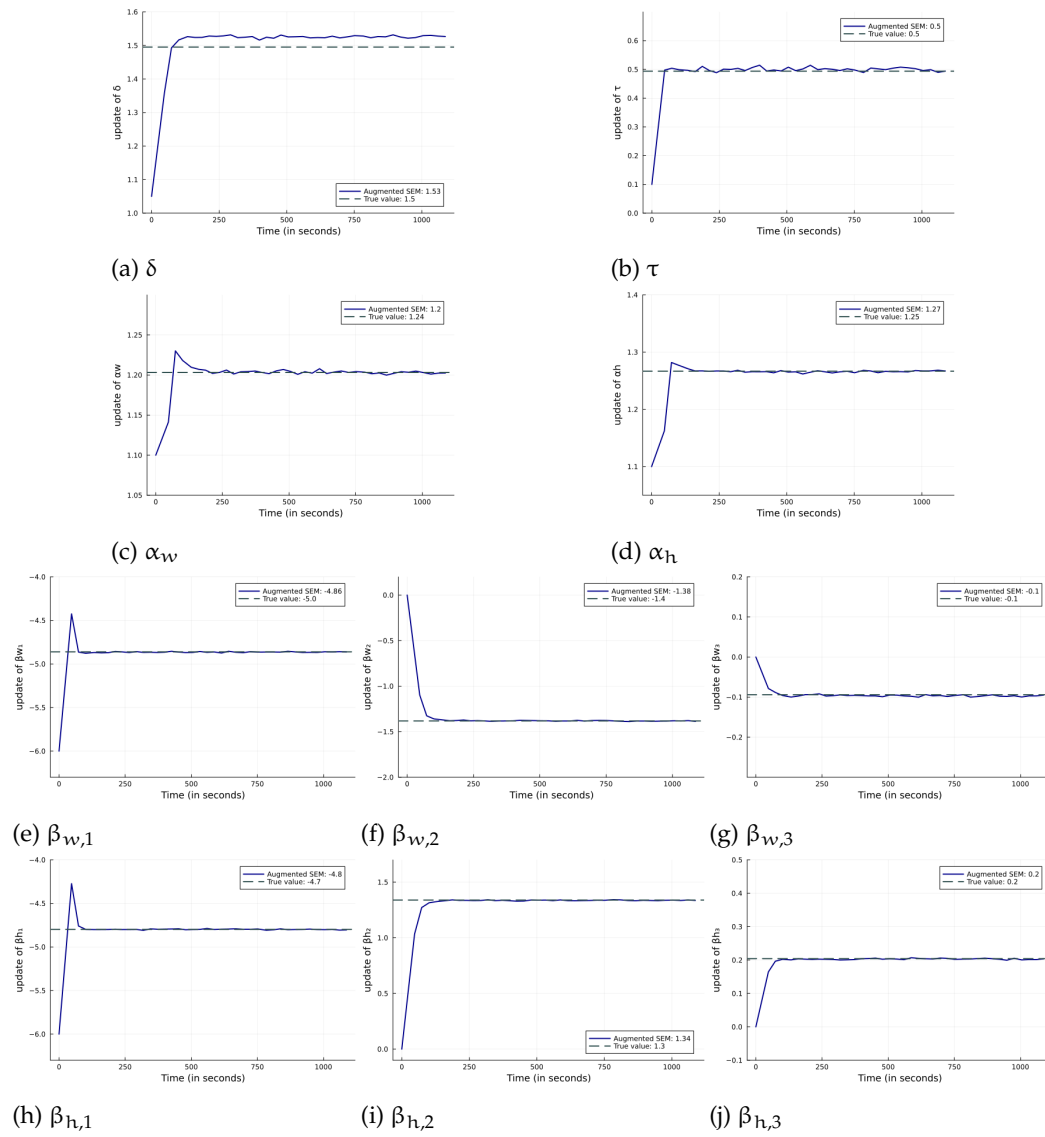
**FIGURE C2** Histograms of the baseline SEM estimates,  $N = 10,000$   
 2 Note: Each panel shows the distribution of baseline SEM estimates across 40 simulations. The dark green vertical line indicates the true parameter value.



**FIGURE C3** Histograms of the baseline and augmented SEM estimates,  $N = 1,000$   
 2 Note: Each panel shows the distributions of the baseline (blue) and augmented SEM (orange) estimates across 40 simulations. The dark green vertical line indicates the true parameter value.



**FIGURE C4** The augmented SEM iterations by computing time from a random initial guess,  $N = 1,000$   
 2 Note: X-axis is the cumulative computing time (seconds) for completing 100 iterations. Blue solid line represents each of the  $M$ -step updates (with extra step). The green dashed line represents the true value.



**FIGURE C5** The augmented SEM iterations by computing time from a random initial guess,  $N = 10,000$   
 2 Note: X-axis is the cumulative computing time (seconds) for completing 40 iterations. Blue solid line represents each of the  $M$ -step update (with extra step). The green dashed line represents the true value.

## D | SUPPLEMENTARY ANALYSES

In this section, we present supplementary analyses related to the estimation of the HDP model, including summary statistics, model fitting, and computing time.

### D.1 | Summary statistics

Here we report summary statistics of the main variables used in the analysis. Table D1 presents sample means of key model variables by gender and censoring status. Censored individuals are on average slightly younger than uncensored individuals.

Table D2 reports the distribution of censoring outcomes by region. A substantial fraction of couples are censored—neither partner retires within the survey period—while only a small fraction, ranging from 8% to 28% across regions, consists of couples in which both partners retire.

Variable	Wives			Husbands		
	Mean (all)	Mean (uncensored)	Mean (censored)	Mean (all)	Mean (uncensored)	Mean (censored)
Age <sub>t0</sub>	56.34	57.67	55.86	58.61	59.417	58.125
Age gap	2.27	1.115	2.68	2.27	2.92	1.882
Primary edu	0.06	0.064	0.059	0.063	0.074	0.057
Tertiary edu	0.354	0.327	0.363	0.338	0.313	0.353
Poor health	0.186	0.163	0.193	0.188	0.166	0.201
Number of grandchildren	1.216	1.497	1.117	1.216	1.331	1.147
Number of children	2.166	2.159	2.168	2.166	2.187	2.153
HH wealth	91246.18	129386.8	77807.91	91246.18	113915.9	77596.4
Southern Europe	0.155	0.074	0.184	0.155	0.113	0.18
Eastern Europe	0.254	0.259	0.252	0.254	0.206	0.283
Western Europe	0.386	0.485	0.351	0.386	0.52	0.305
Northern Europe	0.203	0.181	0.211	0.203	0.159	0.23
N	1969	513	1456	1969	740	1229

**TABLE D1** Summary Statistics of Variables

*Notes: The table reports summary statistics of the data.*

Region	# couples	Both censored	One member	Both retired
Southern Europe	306	0.69	0.23	0.08
Eastern Europe	501	0.60	0.23	0.17
Western Europe	761	0.45	0.27	0.28
Northern Europe	401	0.65	0.18	0.17

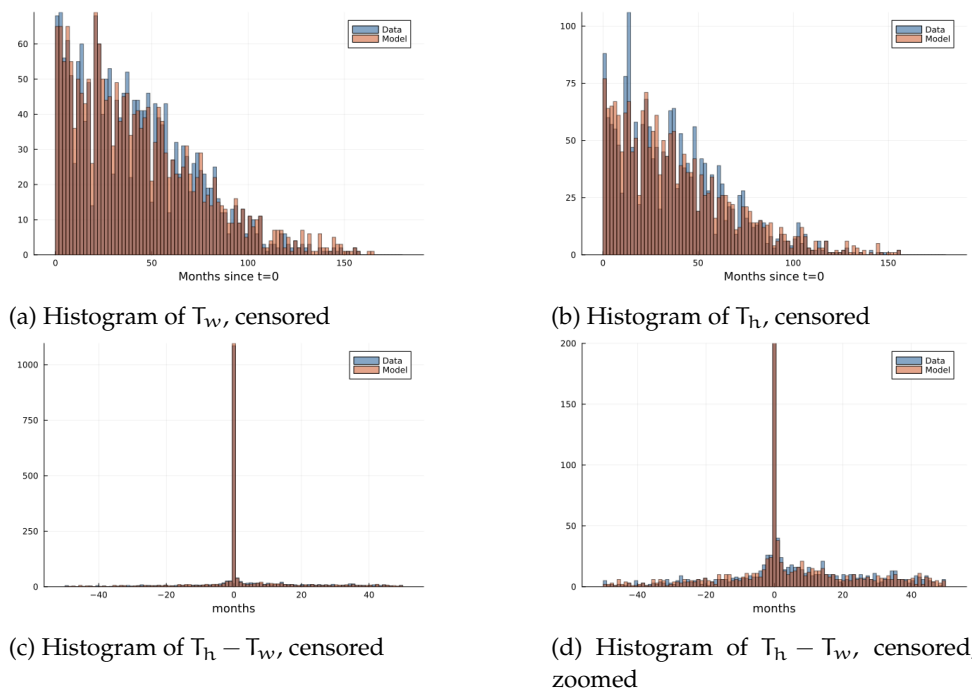
**TABLE D2** Censoring/retirement composition by region

*2 Notes: The table reports the number of couples from each region and the percentage of couples in each retirement status category: both censored, one member censored, and both retired.*

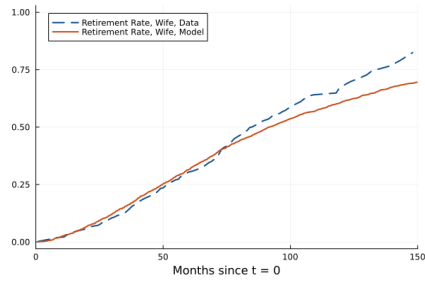
## D.2 | Model fit

In this subsection, we present the results of the model fitting. Specifically, Figure D1 shows the distributions of the censored retirement period  $T$  from both the data (blue) and the simulations generated by the estimated model (orange). The upper two panels present histograms of the censored retirement periods for wives and husbands, respectively, while the lower two panels display the histogram of  $T_h - T_w$ , which contains information on joint retirement patterns. Overall, the estimated model captures the data patterns very well.

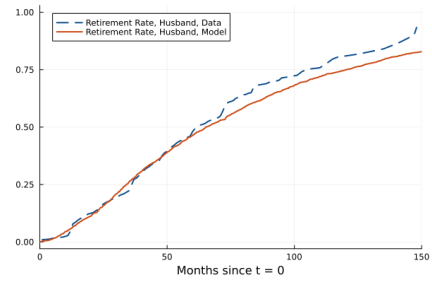
Additionally, Figure D2 shows Kaplan-Meier-based failure functions for the retirement durations of wives and husbands, comparing the empirical data with simulations from the estimated model. Overall, the model reproduces the observed retirement patterns well, although the simulated curves exhibit slightly lower failure rates in the upper tail ( $T > 100$ ), indicating that individuals in the model tend to retire somewhat later than in the data.



**FIGURE D1** Model fit: retirement durations  
 2 Notes: Figures (a-b) compare the histograms of censored retirement time (blue) with the simulated censored retirement time from the estimated model (orange) for wives and husbands, respectively. Figure (c) compares the histograms of the difference in censored retirement time between husbands and wives using data and simulations from the estimated model, and Figure (d) is a zoomed-in version of Figure (c).



(a) Retirement durations, Kaplan Meier, wife



(b) Retirement durations, Kaplan Meier, husband

**FIGURE D2** Model fit: retirement durations

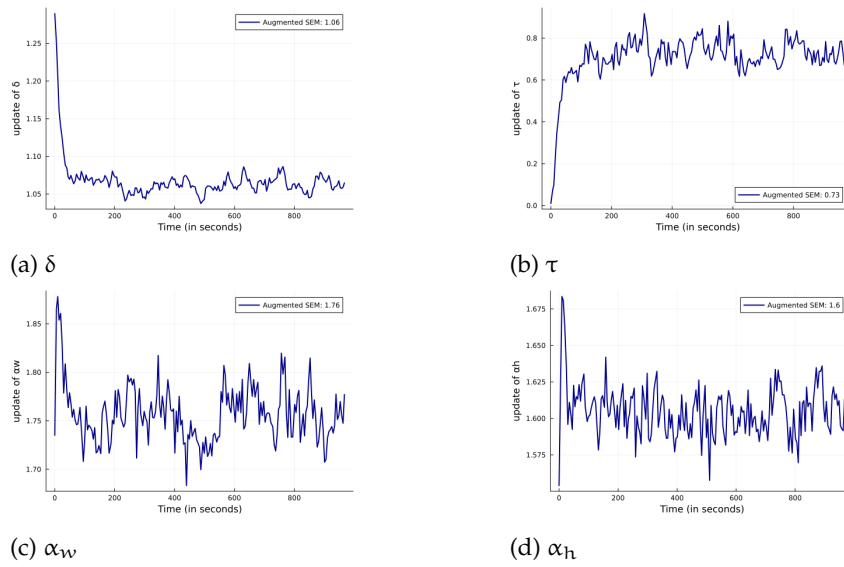
2 Notes: Figures (a-b) compare the estimated retirement durations using Kaplan-Meier plots based on the data (blue dashed line) and simulations from the estimated model (orange solid line) for wives and husbands, respectively.

### D.3 | Estimation results and computing time

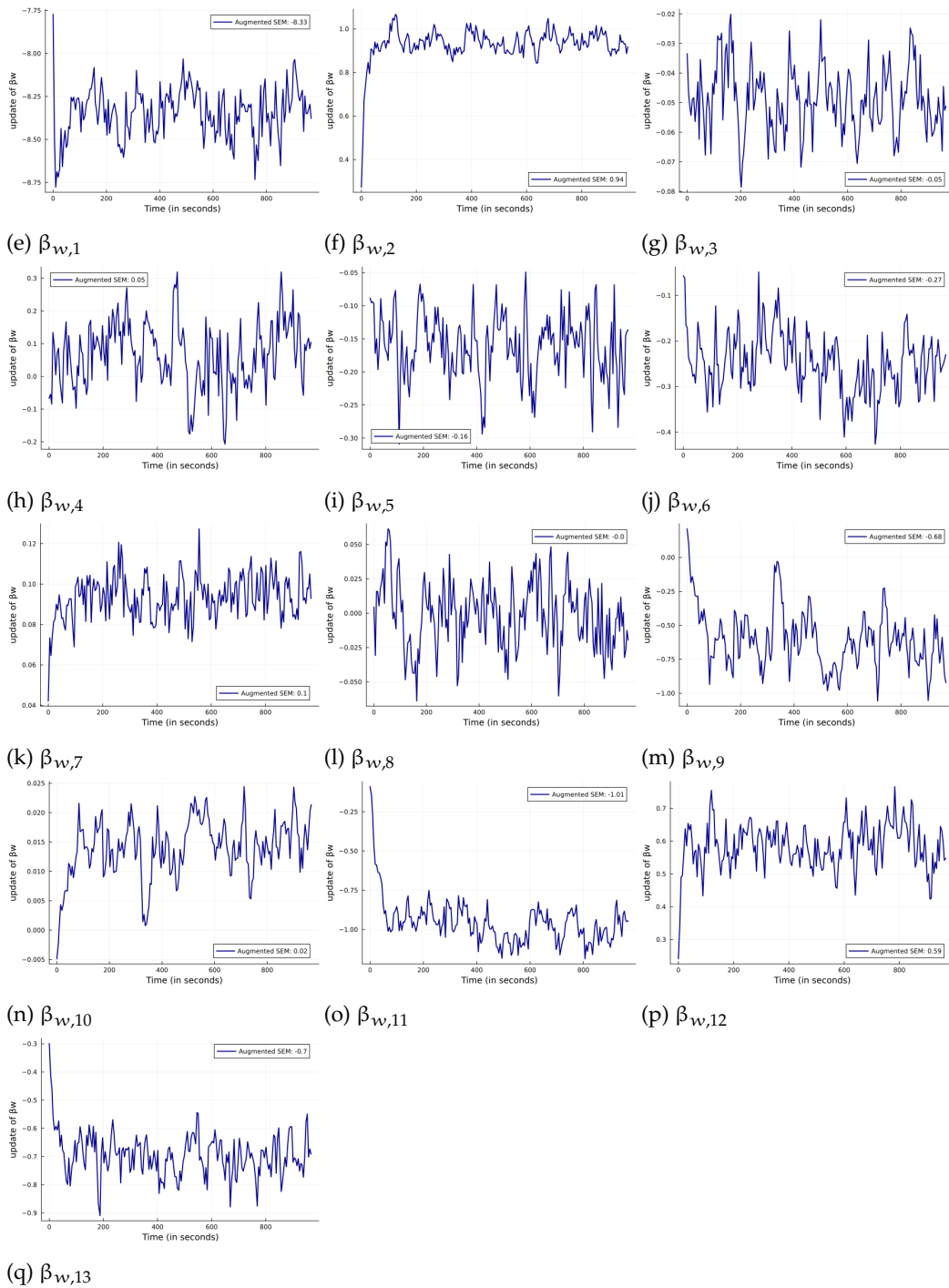
Figure D3 presents the SEM iterations by computing time. The figure shows the cumulative time for running 200 iterations — in total it takes around 15 minutes to complete 200 iterations. Moreover, we can see that the M-step updates converges very quickly, within a few iterations, requiring 1 – 2 minutes. The final estimates are the average of the last 100 iterations.

#### D.4 | Estimation results and computing time

Figure D3 presents the SEM iterations by computing time. The figure shows the cumulative time for running 200 iterations — in total it takes around 15 minutes to complete 200 iterations. Moreover, we can see that the M-step updates converges very quickly, within a few iterations, requiring 1 – 2 minutes. The final estimates are the average of the last 100 iterations.



**FIGURE D3** Augmented SEM iterations by computing time (part 1)  
 2 Note: X-axis is the cumulative computing time (seconds) for completing 200 iterations. Blue solid line represents each of the M-step update (with extra step). We average the last 100 iterations as estimates.



**FIGURE D3** Continued. Results for wives 1.8 Note: X-axis is the cumulative computing time (seconds) for completing 200 iterations. Blue solid line represents each of the  $M$ -step update (with extra step). We average the last 100 iterations as estimates.