

# The informativeness of estimation moments

Bo Honoré<sup>1,2</sup> | Thomas Jørgensen<sup>3</sup> | Áureo de Paula<sup>4</sup>

<sup>1</sup>Department of Economics, Princeton University, Princeton, New Jersey

<sup>2</sup>Dale T. Mortensen Centre at the University of Aarhus, Aarhus, Denmark

<sup>3</sup>Center for Economic Behavior and Inequality (CEBI), Department of Economics, University of Copenhagen, Copenhagen, Denmark

<sup>4</sup>Department of Economics, University College London, CeMMAP and IFS, London, UK

## Correspondence

Áureo de Paula, Department of Economics, University College London, CeMMAP and IFS, London WC1E 6BT, UK.

Email: a.paula@ucl.ac.uk

## Summary

This paper introduces measures for how each moment contributes to the precision of parameter estimates in generalized method of moments settings. For example, one of the measures asks what would happen to the variance of the parameter estimates if a particular moment was dropped from the estimation. The measures are all easy to compute. We illustrate the usefulness of the measures through two simple examples as well as an application to a model of joint retirement planning of couples. We estimate the model using the British Household Panel Survey, and we find evidence of complementarities in leisure. Our sensitivity measures illustrate that the estimate of the complementarity is primarily informed by the distribution of differences in planned retirement dates. The estimated econometric model can be interpreted as a bivariate ordered-choice model that allows for simultaneity. This makes the model potentially useful in other applications.

## 1 | INTRODUCTION

Indirect inference and other nonlinear generalized method of moments (GMM) estimators are used extensively in empirical research. These estimators are, however, sometimes seen as black boxes. It can be difficult to understand exactly what features of the data are informative about which parameters, and how sensitive parameter estimates are to moments included in the objective function.

In this paper, we provide simple and easy-to-compute measures that can indicate how altering the moments used in estimation affects the precision of parameter estimates. Informally, we think of these as measures of how informative each moment is about a particular parameter. More precisely, we provide measures of the effect on asymptotic standard errors from (i) a marginal increase in the noise associated with a moment, (ii) completely removing a (set of) moments from estimation, and (iii) a marginal increase in the weight put on a moment.

The measures are derived from the asymptotic distribution of the class of GMM-type estimators considered here and are, for the most part, based on derivatives of the asymptotic covariance matrix. The measures are almost costless to calculate because most of the required quantities are already constructed when calculating asymptotic standard errors. Furthermore, the measures have straightforward interpretations if scaled in a meaningful way.

There is a growing literature investigating sensitivity of estimators in economics. Recently, for example, Andrews, Gentzkow, and Shapiro (2017) proposed a measure to inform researchers on the sensitivity of the asymptotic bias in estimators to misspecification of moments included in the estimation function. We note that their measure is also related to the change in the asymptotic variance from a marginal change in the included moments, which inspired our proposed alternative measures. While we focus on the precision of the parameter estimates, more recently Armstrong and Kolesár (2018) and Bonhomme and Weidner (2018) also studied local misspecification. Christensen and Connault (2019) studied global misspecification.

-----  
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. Journal of Applied Econometrics published by John Wiley & Sons Ltd.

We illustrate the applicability of our measures through two simple examples and an empirical application. The two examples are a binary outcome probit model and a proportional hazards Weibull duration model with time-varying covariates. The application is a simple structural model of joint retirement planning of dual-earner households. The model is founded in utility maximization with household bargaining, but can also be interpreted as a bivariate ordered-choice model that allows for simultaneity. The parameters of the model are most easily estimated by indirect inference, but the complexity of the model makes it difficult to understand the link between the data and the parameter estimates.

While a growing empirical literature has established that dual-earner households tend to retire simultaneously or in quick succession in age,<sup>1</sup> the empirical evidence of joint retirement *planning* of couples is much more scarce and with ambiguous findings.<sup>2</sup> We contribute to this literature by estimating a structural model of dual-earner retirement planning using indirect inference and prospective retirement planning questions in the British Household Panel Survey (BHPS). Our estimation results support the notion of leisure complementarities in retirement. Our proposed sensitivity measures confirm the intuition that the parameter estimate measuring leisure complementarities in the model is sensitive to the distribution of the difference in the year of planned retirement between household members.

The remaining paper is organized as follows. In Section 2 we present the sensitivity measures and show examples of their use in Section 3. In Section 4 we apply our measures to a novel model of dual-earner retirement planning before concluding with final remarks in Section 5.

## 2 | FRAMEWORK AND SENSITIVITY MEASURES

Indirect inference and other nonlinear GMM estimators are sometimes seen as black boxes where it can be difficult to understand exactly what features of the data are informative about which parameters. In this section, we review and introduce a number of measures that are meant to provide information about this.

To fix ideas, consider a set of moment conditions  $E[f(x_i, \theta_0)] = 0$ , where  $x_i$  is data for observation  $i$  and it is assumed that this defines a unique  $\theta_0$ . The GMM estimator of  $\theta_0$  is  $\hat{\theta} = \arg \min_{\theta} \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta) \right)$ , where  $W_n$  is a symmetric, positive definite matrix. While some of the measures below also apply to just-identified models, we focus here on overidentified models where the number of moments are larger than the number of parameters in  $\theta$ , and the weighting matrix thus plays a role.

Subject to the standard regularity conditions, the derivation of the asymptotic distribution of  $\hat{\theta}$  gives

$$\hat{\theta} = \theta_0 - (G'WG)^{-1}G'W \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0) \right) + o_p(n^{-1/2}), \quad (1)$$

where  $G = E \left[ \frac{\partial f(x_i, \theta_0)}{\partial \theta} \right]$  and  $W$  is the limit of  $W_n$ . See Hansen (1982). The limiting distribution of the GMM estimator is

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = (G'WG)^{-1}G'WSWG(G'WG)^{-1}$$

and  $S = V[f(x_i, \theta_0)]$  under random sampling. If we use the optimal weighting matrix,  $W = S^{-1}$ , the asymptotic covariance collapses to

$$\Sigma_{\text{opt}} = (G'S^{-1}G)^{-1}.$$

Intuitively, when there is little sampling variability in the moment functions,  $f$ ,  $S$  will be small.  $G$  is larger if the moment condition is more sensitive to perturbations in the parameter. Both of these contribute to the precision of the estimates, as the proposed measures highlight.

<sup>1</sup>See, for example, Hurd (1990), Blau (1998), Gustman and Steinmeier (2000, 2004), Coile (2004), An, Christensen, and Gupta (2004), Jia (2005), Blau and Gilleskie (2006), van der Klaauw and Wolpin (2008), Banks, Blundell, and Casanova (2010), Casanova (2010), and Honoré & de Paula, (2018).

<sup>2</sup>See Pienta and Hayward (2002), Moen, Huang, Plassmann, and Dentinger (2006), and de Grip, Fouarge, and Montizaan (2013).

Andrews et al. (2017) proposed the sensitivity measure

$$M_1 = -(G'WG)^{-1}G'W.$$

It is clear from Equation 1 that  $M_1$  provides the mapping from moment misspecification of the type  $E[f(x_i, \theta_0)] = \rho \neq 0$  into parameter biases for small  $\rho$ . Alternatively, by noting that  $\Sigma = M_1SM_1'$ ,  $M_1$  tells us how additional noise in each of the sample moments  $\frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0)$  would result in additional noise in each element of  $\hat{\theta}$ . This is what motivates our alternative measures that address the sensitivity of estimation *precision* to each moment.

The proposed measures are intended to complement the measure of sensitivity to misspecification proposed by Andrews et al. (2017). Like  $M_1$ , our measures are matrices where the  $(j, k)$ th element provides an answer to how the precision of the  $j$ th element of  $\hat{\theta}$  depends on the  $k$ th moment.

Our first measure asks the hypothetical question: How much precision would we lose if the  $k$ th moment is subject to a little additional noise? This measure is formally defined as

$$M_{2,k} \equiv \frac{\partial \Sigma_{\text{opt}}}{\partial S^{(kk)}} = \Sigma_{\text{opt}}(G'S^{-1}O_{kk}S^{-1}G)\Sigma_{\text{opt}},$$

where  $O_{kk}$  is a matrix with 1 in the  $(k, k)$  element and zero elsewhere. This measure assumes that the optimal weighting matrix is used and updated. Alternatively, we could ask the same question keeping the (possibly nonoptimal) weighting matrix unchanged. This measure is

$$M_{3,k} \equiv \frac{\partial \Sigma}{\partial S^{(kk)}} = (G'WG)^{-1}G'WO_{kk}WG(G'WG)^{-1} = M_1O_{kk}M_1'.$$

The difference between  $M_{2,k}$  and  $M_{3,k}$  is that the former evaluates the potential information in each moment, while the latter evaluates the information actually used in the estimation. With efficient GMM (so  $W = S^{-1}$ ),  $M_3$  equals  $M_2$ . This is also true in the just-identified case where the number of moments equals the number of parameters to be estimated.

Related to  $M_{2,k}$ , we could consider the change in the asymptotic variance from completely excluding the  $k$ th moment:

$$M_{4,k} \equiv \tilde{\Sigma}_k - \Sigma,$$

where

$$\begin{aligned} \tilde{\Sigma}_k &= (G'\tilde{W}_kG)^{-1}G'\tilde{W}_kS\tilde{W}_kG(G'\tilde{W}_kG)^{-1}, \\ \tilde{W}_k &= W \odot (I_k I_k'). \end{aligned}$$

Here  $\odot$  denotes element-wise multiplication and  $I_k$  is a  $J \times 1$  vector with ones in all elements except the  $k$ th element, which is zero.  $M_{4,k}$  leaves the weighting matrix on the remaining moments unchanged after we have excluded the  $k$ th moment.

We note that this measure assumes that the parameter vector is identified after the  $k$ th moment has been excluded. Specifically,  $(G'\tilde{W}_kG)$  needs to have full rank. Importantly, this means that the original model has to be overidentified in the sense that it has more moments than parameters. In practice,  $G$  has to be estimated, and violations of the full rank assumption will result in  $(\hat{G}'\tilde{W}_k\hat{G})$  being close to singular. Extremely large values in the estimate of  $M_{4,k}$  therefore suggest that the model is not point-identified when the  $k$ th moment is excluded. This can happen even if the original model was overidentified.

Alternatively, one could also consider measures that adjust the weighting matrix. For example, one could consider a measure that compares the precision of the optimal GMM estimator that uses all moments to the optimal GMM estimator that excludes that  $k$ th moment:

$$M_{5,k} = (G'_{-k}S_{-k}^{-1}G_{-k})^{-1} - (G'S^{-1}G)^{-1},$$

where  $G_{-k}$  is the same as matrix  $G$  except that the  $k$ th row has been removed, and  $S_{-k}$  is  $S$  with the  $k$ th row and column removed. This measure also assumes that the parameter vector is identified after the  $k$ th moment has been excluded, and it implicitly assumes that the original number of moment conditions exceeds the number of parameters to be estimated.

$M_4$  and  $M_5$  can also be used to gauge the sensitivity of the estimator to a set of moments. This is potentially useful in cases where one can group moments in some natural way. One can then address the question of how much of the precision in an estimator would be lost if one did not use one of the groups of moments. For example, (Gayle & Shephard, 2019) talks about five sets of moments (in their online appendix), and Honoré and de Paula (2018) get their moments from the estimation of four different auxiliary reduced form models. A reparametrization of those reduced-form models would lead to moment conditions that are (asymptotically) linear combinations of the original moment conditions. In that case, it might be useful to construct a measure that reflects giving zero weight to all the moments that come from a specific auxiliary model. This approach would be application specific, and we therefore do not pursue it in this paper.

Our final measure addresses the question: How would the precision of our estimates change if we slightly increased the weight put on the  $k$ th moment? This measure is formally defined as the derivative

$$M_{6,k} \equiv \frac{\partial \Sigma}{\partial W^{(k,k)}} = -(G'WG)^{-1}(G'O_{kk}G)\Sigma + (G'WG)^{-1}G'O_{kk}SWG(G'WG)^{-1} \\ + (G'WG)^{-1}G'WSO_{kk}G(G'WG)^{-1} - \Sigma(G'O_{kk}G)(G'WG)^{-1}.$$

We do not think of  $M_{6,k}$  as a measure of moment sensitivity, but rather as a measure of how close the chosen weighting matrix is to being optimal.  $M_{6,k}$  will be 0 when  $W$  is the optimal weighting matrix. It will also be 0 in the just-identified case, where the number of moments equals the number of parameters to be estimated.

These measures are not invariant to the scale of the included moments in  $f(\cdot)$ . One approach, which we take, is to report scaled measures. Concretely, we report the sensitivity of the  $j$ th parameter to the  $k$ th moment as

$$\mathcal{E}_2^{(j,k)} = M_2^{(j,k)} \frac{S^{(k,k)}}{\Sigma^{(j,j)}}, \\ \mathcal{E}_3^{(j,k)} = M_3^{(j,k)} \frac{S^{(k,k)}}{\Sigma^{(j,j)}}, \\ \mathcal{E}_4^{(j,k)} = M_4^{(j,k)} \frac{1}{\Sigma^{(j,j)}}, \\ \mathcal{E}_5^{(j,k)} = M_5^{(j,k)} \frac{1}{\Sigma_{opt}^{(j,j)}}, \\ \mathcal{E}_6^{(j,k)} = M_6^{(j,k)} \frac{W^{(k,k)}}{\Sigma^{(j,j)}},$$

where  $M_2^{(j,k)}$  refers to the  $j$ th diagonal element of  $M_{2,k}$  and similarly for  $M_3^{(j,k)}$ ,  $M_4^{(j,k)}$ ,  $M_5^{(j,k)}$  and  $M_6^{(j,k)}$ . Note that  $\mathcal{E}_2^{(j,k)}$ ,  $\mathcal{E}_3^{(j,k)}$  and  $\mathcal{E}_6^{(j,k)}$  are elasticities, whereas  $\mathcal{E}_4^{(j,k)}$  and  $\mathcal{E}_5^{(j,k)}$  are the relative changes in the asymptotic variance compared to the baseline with all moments included.

### 3 | EXAMPLES

In this section, we illustrate the use of our proposed measures through two concrete examples. The first example is a simple binary-choice probit model and the second example is a proportional hazards duration model. The first example is chosen because it is a case where one would have a strong prior about which moments matter. The second example, on the other hand, is complicated enough that this is not obvious.

For both examples, we use both the optimal weighting matrix and a diagonal weighting matrix with the inverse of the moment variances on the diagonal. We chose the latter nonoptimal weighting matrix because it is very common in empirical applications.<sup>3</sup>

<sup>3</sup>There are many examples of this. They include Eisenhauer, Heckman, and Mosso (2015) and Gayle and Shephard (2019), to name two. The motivation stems from Altonji and Segal (1996), who show that the optimal weighting matrix can have quite poor finite sample properties. They suggest equally weighted moments (i.e.,  $W = I$ ) as an alternative. Of course, using equal weights will not be invariant to changes in units (or other rescaling), which explains the practice we have adopted.

**TABLE 1** Sensitivity measures, probit model, optimal weighting

		Moment					
		$\mathbb{E}[e]$	$\mathbb{E}[ex_1]$	$\mathbb{E}[ex_2]$	$\mathbb{E}[ex_1^2]$	$\mathbb{E}[ex_1x_2]$	$\mathbb{E}[ex_2^2]$
$M_1$	$\beta_0$	4.261	1.475	1.469	0.192	0.378	0.184
	$\beta_1$	1.190	6.570	-1.286	0.223	0.141	-0.069
	$\beta_1$	1.193	-1.286	6.567	-0.073	0.152	0.214
$\mathcal{E}_2$	$\beta_0$	1.104	0.088	0.087	0.003	0.004	0.003
	$\beta_1$	0.060	1.207	0.046	0.003	0.000	0.000
	$\beta_1$	0.060	0.046	1.205	0.000	0.000	0.003
$\mathcal{E}_3$	$\beta_0$	1.104	0.088	0.087	0.003	0.004	0.003
	$\beta_1$	0.060	1.207	0.046	0.003	0.000	0.000
	$\beta_1$	0.060	0.046	1.205	0.000	0.000	0.003
$\mathcal{E}_4$	$\beta_0$	1.206	0.292	0.291	0.005	0.003	0.005
	$\beta_1$	0.065	4.014	0.155	0.004	0.006	0.000
	$\beta_1$	0.065	0.153	4.034	0.000	0.006	0.004
$\mathcal{E}_5$	$\beta_0$	1.203	0.292	0.291	0.001	0.003	0.001
	$\beta_1$	0.065	4.014	0.155	0.001	0.000	0.000
	$\beta_1$	0.065	0.153	4.034	0.000	0.000	0.001
$\mathcal{E}_6$	$\beta_0$	0.000	0.000	0.000	0.000	0.000	0.000
	$\beta_1$	0.000	0.000	0.000	0.000	0.000	0.000
	$\beta_1$	0.000	0.000	0.000	0.000	0.000	0.000

Note. Simulations based on  $10^7$  observations.

### 3.1 | Example 1: Method of moments estimation of a probit model

We first consider a simple probit model:

$$y_i = \begin{cases} 0 & \text{if } y_i^* > 0 \\ 1 & \text{else,} \end{cases}$$

$$y_i^* = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i,$$

where  $(x_{1,i}, x_{2,i})$  has a bivariate normal distribution with means equal to 0, variance 1, and correlation 0.5.  $\varepsilon_i$  is independent of  $(x_{1,i}, x_{2,i})$  and distributed according to a standard normal. We set  $\beta_0 = \beta_1 = \beta_2 = 1/\sqrt{3}$ . This makes  $V[\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}] = 1$  and  $P(y_i = 1) = 0.66$ .<sup>4</sup>

We consider the asymptotic distribution of a moment-based estimator of  $\theta_0 = (\beta_0, \beta_1, \beta_2)$  solving

$$\hat{\theta} = \arg \min_{\theta} g(\theta)' W g(\theta),$$

where we use the six moments:

$$(E[e(\theta)] \ E[e(\theta)x_1] \ E[e(\theta)x_2] \ E[e(\theta)x_1^2] \ E[e(\theta)x_1x_2] \ E[e(\theta)x_2^2])'$$

and  $e_i(\theta) = y_i - \Phi(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})$ . In the corresponding logit model, the first three moments correspond to the first-order conditions for maximum likelihood estimation. Although they are formally different, the logit and probit models are quite similar. We therefore expect the first three moments to be the most informative about  $\theta_0$ . Moreover, we expect the first moment to be the most important for determining  $\hat{\beta}_0$ , and the second and third for determining  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , respectively.

Table 1 shows results using the optimal weighting matrix, and Table 2 shows results using the diagonal weighting matrix with the inverse of the moment variances on the diagonal.<sup>5</sup> We think of the latter as a practical alternative to the efficient weighting matrix.

<sup>4</sup>We also supply Python code to illustrate our approach.

<sup>5</sup>We illustrate the proposed sensitivity measures through Monte Carlo simulation of the expected values using  $10^7$  simulated observations.

**TABLE 2** Sensitivity measures, probit model, diagonal weighting

		Moment					
		$\mathbb{E}[e]$	$\mathbb{E}[ex_1]$	$\mathbb{E}[ex_2]$	$\mathbb{E}[ex_1^2]$	$\mathbb{E}[ex_1x_2]$	$\mathbb{E}[ex_2^2]$
$M_1$	$\beta_0$	3.374	1.633	1.630	1.036	-0.681	1.035
	$\beta_1$	1.354	5.656	-1.185	-0.853	-1.360	0.882
	$\beta_1$	1.351	-1.185	5.658	0.881	-1.360	-0.851
$\mathcal{E}_2$	$\beta_0$	1.104	0.088	0.087	0.003	0.004	0.003
	$\beta_1$	0.060	1.207	0.046	0.003	0.000	0.000
	$\beta_1$	0.060	0.046	1.205	0.000	0.000	0.003
$\mathcal{E}_3$	$\beta_0$	0.651	0.101	0.101	0.080	0.013	0.080
	$\beta_1$	0.071	0.817	0.036	0.037	0.034	0.039
	$\beta_1$	0.070	0.036	0.817	0.039	0.034	0.037
$\mathcal{E}_4$	$\beta_0$	1.076	0.341	0.340	-0.010	-0.011	-0.011
	$\beta_1$	0.042	3.783	0.116	-0.038	-0.031	-0.028
	$\beta_1$	0.042	0.114	3.802	-0.028	-0.032	-0.038
$\mathcal{E}_5$	$\beta_0$	1.203	0.292	0.291	0.001	0.003	0.001
	$\beta_1$	0.065	4.014	0.155	0.001	0.000	0.000
	$\beta_1$	0.065	0.153	4.034	0.000	0.000	0.001
$\mathcal{E}_6$	$\beta_0$	-0.101	0.002	0.002	0.040	0.017	0.041
	$\beta_1$	0.011	-0.142	0.002	0.044	0.048	0.037
	$\beta_1$	0.011	0.002	-0.142	0.037	0.048	0.044

Note. Simulations based on  $10^7$  observations.

It is clear from Table 1 that the first three moments are indeed the most informative about  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ , respectively. As mentioned, this is expected since these moments would be the first-order conditions for maximum likelihood estimation of a logit model.

The elements in the last three columns of  $M_1$  in Table 1 are much smaller than those in the first three. This suggests that the optimal GMM estimator is much less sensitive to misspecification of the last three moments than to misspecification of the first three moments. The reason is that the first three moments get almost all the weight (in the corresponding logit model, they would literally get all the weight). As expected, this is less pronounced in Table 2. The values of  $\mathcal{E}_2$  in Tables 1 and 2 confirm that the efficient GMM estimator of  $\theta_0$  is driven by the first three moments.<sup>6</sup> Adding noise to the last three moments has essentially no effect on the precision of the optimal GMM estimator of  $\theta_0$ , whereas adding noise to the first three elements can have a big effect. The values of  $\mathcal{E}_3$  in Table 2 illustrate that the precision of the nonoptimal GMM estimator is less sensitive to noise in the last three moments (because they get relatively less weight) and more sensitive to adding noise to the first three moments (because they get relatively more weight).

Next,  $\mathcal{E}_4$  and  $\mathcal{E}_5$  suggest that leaving out, for example, the second moment would increase the asymptotic variance of both the efficient and the inefficient GMM estimator of  $\beta_1$  by around 400%. This confirms that  $E[ex_1]$  is instrumental for precise estimation of  $\beta_1$ .

The final measure,  $\mathcal{E}_6$  in Table 1, is 0 by construction. Since we are using the weighting matrix that minimizes the variance of the estimator of each element of  $\theta$ , the derivative of the variance with respect to the elements of the weighting matrix must be 0.  $\mathcal{E}_6$  in Table 2 shows that in this case the diagonal weighting matrix with the inverse of the moment variances on the diagonal puts too little weight on the first three moments.

### 3.2 | Example 2: Duration model

The probit example in Section 3.1 was chosen because it is an example where we have good prior intuition about which moments matter for what parameter. We now turn to an example where this is much less obvious.

Consider a duration,  $T$ , that follows a mixed proportional hazard model with time-varying covariates and a Weibull as the baseline hazard:

$$h(t) = \alpha t^{\alpha-1} \exp(x'(t)\beta) \eta,$$

<sup>6</sup> $\mathcal{E}_2$  in Tables 1 and 2 differ only because of simulation error.

where  $\alpha$  is the scale parameter that captures duration dependence and  $x'(t)\beta$  is the effect of the time-varying explanatory variables. An example of a two-dimensional time-varying set of explanatory variables could be

$$x(t) = \begin{cases} (x_{1,1}, x_{2,1}) & \text{if } t < t_1 \\ (x_{1,2}, x_{2,2}) & \text{if } t_1 \leq t \leq t_2 \\ \vdots & \vdots \\ (x_{1,k}, x_{2,k}) & \text{if } t_{k-1} \leq t. \end{cases}$$

Finally,  $\eta$  captures unobserved heterogeneity. Except for moment assumptions, no assumptions are made on the distribution of  $\eta$ .

We then have the survival function for  $T$ :

$$S(t|x(\cdot), \eta) = \exp\left(-\eta \int_0^t \alpha s^{\alpha-1} \exp[x'(s)\beta] ds\right).$$

Since

$$S(T|x(\cdot), \eta) \sim U(0, 1),$$

we have

$$\eta \int_0^T \alpha s^{\alpha-1} \exp[x'(s)\beta] ds \sim \text{Exp}(1), \text{ conditional on } x(\cdot), \eta$$

or

$$\log\left(\int_0^T \alpha s^{\alpha-1} \exp(x'(s)\beta) ds\right) \sim \log(\text{Exp}(1)) - \log(\eta), \text{ conditional on } x(\cdot), \eta. \quad (2)$$

Here,  $\text{Exp}(1)$  denotes an exponentially distributed random variable with mean 1, and  $-\log(\text{Exp}(1))$  follows a standard Gumbel distribution with  $E[-\log(\text{Exp}(1))] = \gamma \approx 0.57721$  (Euler's constant) and  $V[-\log(\text{Exp}(1))] = \pi^2/6$ .

Equation 2 suggests moment conditions of the type

$$E\left[\left(\log\left(\int_0^T \alpha s^{\alpha-1} \exp[x'(s)\beta] ds\right) + \gamma - \beta_0\right) \psi[x(\cdot)]\right] = 0 \quad (3)$$

for functions of the covariates,  $\psi$ . Here,  $\beta_0$  captures the mean of  $-\log(\eta)$ , which is assumed to be finite.

When  $x(t)$  is time invariant, Equation 2 becomes

$$\log(T^\alpha \exp(x'\beta)) \sim \log(\text{Exp}(1)) - \log(\eta)$$

or

$$\log(T) = -x'(\beta/\alpha) + \text{“error.”}$$

In other words, with time-invariant covariates the moments implied by Equation 3 do not identify  $(\beta, \alpha)$ , but only  $\beta/\alpha$ . It turns out that it is possible to estimate  $\alpha$  by other methods (see, e.g., Honoré, 1990), but it is not possible to estimate  $(\beta, \alpha)$  at the usual  $\sqrt{n}$  rate (see Hahn, 1994). This makes it interesting to investigate how precision in estimation of  $(\beta, \alpha)$  depends on the various moments in Equation 3 when  $x$  does contain time-varying covariates.

We consider a data-generating process with one time-invariant and one time-varying covariate. Specifically,  $x(s) = [x_1(s), x_2(s)]$ , where

$$x(s) = \begin{cases} (x_1, x_{21}) & \text{for } s \leq 1 \\ (x_1, x_{22}) & \text{for } 1 < s \leq 2 \\ (x_1, x_{23}) & \text{for } 2 < s, \end{cases}$$

with  $x_1 = Z_1$ ,  $x_{21} = Z_2$ ,  $x_{22} = (x_{21} + Z_3)/\sqrt{2}$  and  $x_{23} = (x_{22} + Z_4)/\sqrt{2}$ .  $Z_1$  through  $Z_4$  follow standard normal distributions. The heterogeneity term,  $\eta$ , follows a log-normal distribution, where the underlying normal has mean 0 and variance  $1/2$ .  $\eta$  is independent of  $x(\cdot)$ . Finally,  $\beta = (-1, 1/\sqrt{2}, 1/\sqrt{2})'$  and  $\alpha = 2$ . With this, the median duration is approximately

**TABLE 3** Sensitivity measures, Weibull model, optimal weighting

		Moment				
		$E[e]$	$E[ex_1]$	$E[ex_{21}]$	$E[ex_{22}]$	$E[ex_{23}]$
$M_1$	$\beta_0$	-0.503	0.001	3.375	-2.053	-1.934
	$\beta_1$	-0.000	-0.500	0.018	-0.015	-0.014
	$\beta_2$	-0.000	0.000	-0.228	-0.251	-0.184
	$\alpha$	-0.019	0.009	24.478	-15.092	-14.181
$\mathcal{E}_2$	$\beta_0$	0.028	0.000	1.282	0.474	0.421
	$\beta_1$	0.000	0.998	0.001	0.001	0.001
	$\beta_2$	0.000	0.000	0.155	0.187	0.100
	$\alpha$	0.000	0.000	1.299	0.494	0.436
$\mathcal{E}_3$	$\beta_0$	0.028	0.000	1.282	0.474	0.421
	$\beta_1$	0.000	0.998	0.001	0.001	0.001
	$\beta_2$	0.000	0.000	0.155	0.187	0.100
	$\alpha$	0.000	0.000	1.299	0.494	0.436
$\mathcal{E}_4$	$\beta_0$	>100 <sup>a</sup>	>100 <sup>a</sup>	4.841	0.196	0.274
	$\beta_1$	0.324 <sup>a</sup>	>100 <sup>a</sup>	0.005	0.000	0.001
	$\beta_2$	0.012 <sup>a</sup>	>100 <sup>a</sup>	0.584	0.077	0.065
	$\alpha$	3.935 <sup>a</sup>	>100 <sup>a</sup>	4.904	0.203	0.284
$\mathcal{E}_5$	$\beta_0$	>100 <sup>a</sup>	>100 <sup>a</sup>	4.841	0.196	0.274
	$\beta_1$	0.324 <sup>a</sup>	>100 <sup>a</sup>	0.005	0.000	0.001
	$\beta_2$	0.012 <sup>a</sup>	>100 <sup>a</sup>	0.584	0.077	0.065
	$\alpha$	3.935 <sup>a</sup>	>100 <sup>a</sup>	4.904	0.203	0.284
$\mathcal{E}_6$	$\beta_0$	0.000	-0.000	0.000	0.000	0.000
	$\beta_1$	-0.000	-0.000	0.000	0.000	0.000
	$\beta_2$	0.000	0.000	-0.000	0.000	0.000
	$\alpha$	-0.000	-0.000	0.000	0.000	0.000

Note. Simulations based on  $10^7$  observations.

<sup>a</sup>As mentioned in the text, large values of  $\mathcal{E}_4$  and  $\mathcal{E}_5$  suggest that the model is not identified after the moment has been removed from estimation.

1.3, approximately 38% of the durations are less than 1, and 29% greater than 2. This design is chosen because it is a simple example with sizable unobserved heterogeneity and duration dependence, and where we expect that the time-varying covariate might have bite. The design is not meant to mimic any realistic empirical example.

We again consider a moment-based estimator of  $\theta = (\beta_0/\alpha, \beta_1/\alpha, \beta_2/\alpha, \alpha)$  solving

$$\hat{\theta} = \arg \min_{\theta} g(\theta)' Wg(\theta),$$

where we use the five moments given by Equation 3 with  $\psi[x(\cdot)] = (1, x_1, x_{21}, x_{22}, x_{23})$ .

The sensitivity measures are given in Tables 3 and 4. In this design, the derivative of the first two moments at the true parameter values are nonzero with respect to  $\theta_0$  and  $\theta_1$ , respectively. The derivatives are 0 with respect to the other parameters. This implies that  $G$  becomes singular when we exclude either of the first two moments. This explains the extreme entries for  $\mathcal{E}_4$  and  $\mathcal{E}_5$  in Tables 3 and 4.

The conclusions from the remaining parts of the sensitivity measures are fairly consistent. Most interestingly, the moments formed on the basis of the time-varying covariates contribute to the identification of  $\alpha$ , while the moment based on the time-invariant covariate does not. This is exactly what the discussion above would predict. Interestingly, the first moment is also important for  $\alpha$ . Presumably, this is because this moment determines the estimate of the mean of the (log of the) unobserved heterogeneity. It is well known in the duration literature that unobserved heterogeneity is poorly distinguished from duration dependence. As a result, we do not consider this surprising.

## 4 | APPLICATION: JOINT RETIREMENT PLANNING

In this section, we apply the proposed sensitivity measures to an extremely simple structural model of the joint retirement planning of dual-earner couples.

**TABLE 4** Sensitivity measures, Weibull model, diagonal weighting

		Moment				
		$E[e]$	$E[ex_1]$	$E[ex_{21}]$	$E[ex_{22}]$	$E[ex_{23}]$
$M_1$	$\beta_0$	-0.503	0.001	3.066	-1.117	-2.679
	$\beta_1$	-0.000	-0.500	0.016	-0.010	-0.019
	$\beta_2$	-0.000	0.000	-0.234	-0.234	-0.197
	$\alpha$	-0.021	0.009	22.219	-8.255	-19.619
$\mathcal{E}_2$	$\beta_0$	0.028	0.000	1.282	0.474	0.421
	$\beta_1$	0.000	0.998	0.001	0.001	0.001
	$\beta_2$	0.000	0.000	0.155	0.187	0.100
	$\alpha$	0.000	0.000	1.299	0.494	0.436
$\mathcal{E}_3$	$\beta_0$	0.027	0.000	1.017	0.135	0.775
	$\beta_1$	0.000	0.998	0.001	0.000	0.001
	$\beta_2$	0.000	0.000	0.162	0.163	0.115
	$\alpha$	0.000	0.000	1.027	0.142	0.800
$\mathcal{E}_4$	$\beta_0$	>100 <sup>a</sup>	>100 <sup>a</sup>	4.612	0.149	0.224
	$\beta_1$	0.323 <sup>a</sup>	>100 <sup>a</sup>	0.005	0.000	0.000
	$\beta_2$	0.011 <sup>a</sup>	>100 <sup>a</sup>	0.583	0.077	0.065
	$\alpha$	3.737 <sup>a</sup>	>100 <sup>a</sup>	4.667	0.155	0.232
$\mathcal{E}_5$	$\beta_0$	>100 <sup>a</sup>	>100 <sup>a</sup>	4.841	0.196	0.274
	$\beta_1$	0.324 <sup>a</sup>	>100 <sup>a</sup>	0.005	0.000	0.001
	$\beta_2$	0.012 <sup>a</sup>	>100 <sup>a</sup>	0.584	0.077	0.065
	$\alpha$	3.935 <sup>a</sup>	>100 <sup>a</sup>	4.904	0.203	0.284
$\mathcal{E}_6$	$\beta_0$	0.000	-0.000	-0.049	-0.054	0.102
	$\beta_1$	0.000	0.000	-0.000	-0.000	0.000
	$\beta_2$	0.000	-0.000	0.002	-0.006	0.004
	$\alpha$	0.000	-0.000	-0.049	-0.056	0.105

Note. Simulations based on  $10^7$  observations.

<sup>a</sup>As mentioned in the text, large values of  $\mathcal{E}_4$  and  $\mathcal{E}_5$  suggest that the model is not identified after the moment has been removed from estimation.

## 4.1 | Data and institutional setting

We use the British Household Panel Survey (BHPS), which is a completed panel of 18 waves collected from 1991 to 2009. In waves 11 and 16 of the BHPS, each adult household member is asked, “Even if this is some time away, at what age do you expect you will retire?” We use this to measure the subjective retirement plans of each spouse.<sup>7</sup> Based on the age at the interview and the expected retirement age, we can calculate the expected retirement year of each household member and use that to investigate joint retirement plans.

Besides retirement plans, we use information in the BHPS on annual labor market income, the number of visits to the general practitioner (GP), subjective expectations about future health status, eligibility for an employer-provided pension scheme (EPP), and whether individuals save any of their income in a private personal pension (PPP).<sup>8</sup> Finally, we define individuals as highly skilled if they have completed the first or second stage of tertiary education (ISCED code 5 or 6).

We use information on households consisting of two opposite-sex household members who are either married or cohabiting, and who meet the following sample selection criteria: (i) both members are between 40 and 59 years old when interviewed; (ii) at least one member is not retired at the time of the interview; and (iii) retirement plans are observed in the age range 50–70 for at least one member not retired at the time of the interview. If a household satisfies the criteria in both waves (11 and 16), we use both survey responses in the analysis. We refer to each household member as husband or wife, although we also include households where couples are cohabiting but not necessarily married.

### 4.1.1 | The state pension age (SPA)

The SPA in the UK is the age at which individuals become eligible to receive state pension from the government. Individuals who have reached SPA and contributed to the scheme for sufficient years are eligible to receive a weekly

<sup>7</sup>The exact formulation in wave 11 is slightly different: “At what age do you expect to retire/will you consider yourself to be retired?”

<sup>8</sup>The EPP includes both defined and contributed benefit (DB and CB) plans and we cannot distinguish between them. Blundell et al. (2004) show, however, that DB plans were most common in the UK in this period.

TABLE 5 Descriptive Statistics

	Mean	SD	Min	Max	Obs.
Age, husband	49.613	5.53	40	59	1,730
Age, wife	48.128	5.34	40	59	1,730
Planned retirement age, husband	62.606	3.87	50	70	1,730
Planned retirement age, wife	60.301	3.72	50	70	1,730
Diff. in planned retirement year (husband – wife)	0.823	5.71	–20	27	1,730
High skilled, husband	0.157	0.36	0	1	1,730
High skilled, wife	0.139	0.35	0	1	1,730
10+ GP visits, husband	0.039	0.19	0	1	1,729
10+ GP visits, wife	0.080	0.27	0	1	1,729
Expect worse health, husband	0.182	0.39	0	1	1,641
Expect worse health, wife	0.115	0.32	0	1	1,645
Labor income (£1,000), husband	25.248	17.12	0	244	1,600
Labor income (£1,000), wife	13.815	10.78	0	109	1,442
Private pension, husband	0.280	0.45	0	1	1,730
Private pension, wife	0.134	0.34	0	1	1,730
Employer pension, husband	0.514	0.50	0	1	1,730
Employer pension, wife	0.466	0.50	0	1	1,730

transfer with no means testing. In 2009, the weekly rate was around £95. See Bozio, Crawford, and Tetlow (2010), Blundell, Meghir, and Smith (2004), and Cribb, Emmerson, and Tetlow (2013) for excellent descriptions of the pension system in the UK.

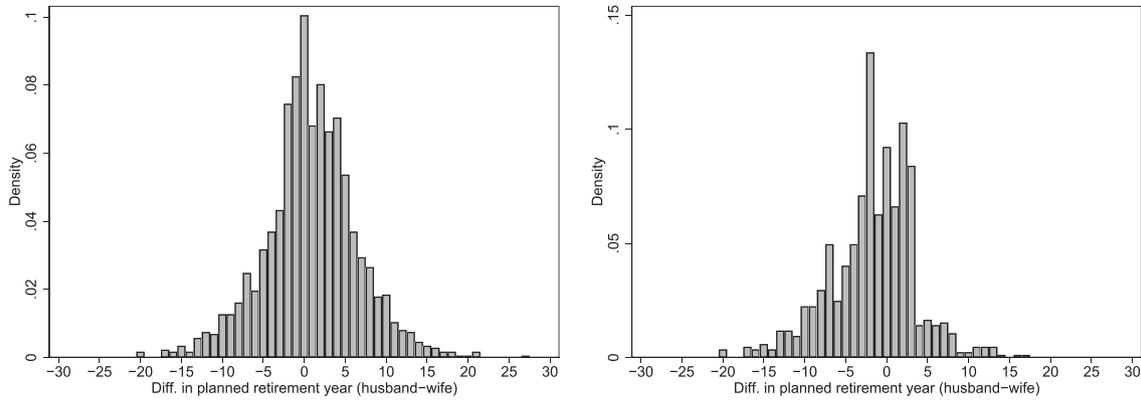
The SPA was 65 for men and 60 for women until the implementation of the Pension Act 1995. The Pension Act 1995 introduced an increase in the SPA of women born after April 6, 1950. While the SPA for men was unaffected, the SPA for women was gradually increased by 1 month every month (by date of birth) until the SPA for women reached 65 for cohorts born later than (including) 1955. See Thurley and Keen (2017) for a comprehensive discussion of the reform.<sup>9</sup> Since this might affect individual expectations, our modeling framework explicitly allows for an effect of the Pension Act 1995 on retirement planning.

#### 4.1.2 | Descriptive statistics

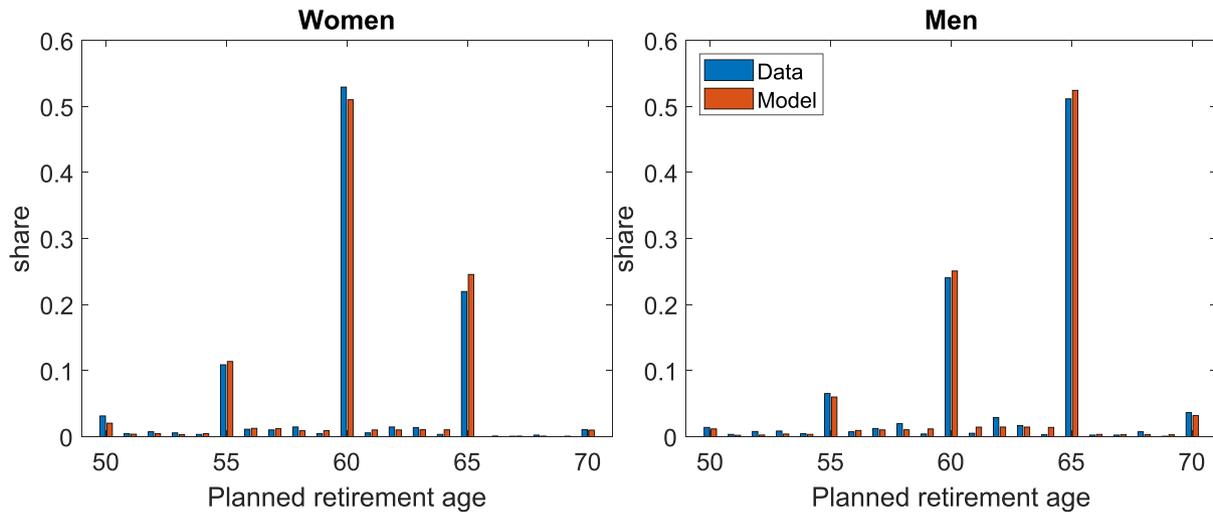
Table 5 reports the descriptive statistics for the variables that we use. All statistics are based on households in which both members are not retired at the time of the interview, which is around 97% of our sample. Husbands in the estimation sample are approximately 1.5 years older than their wives, plan to retire 2 years later than their wives (at age 63 on average), and the average difference in the planned retirement year is approximately 0.83 years. This difference should be viewed in light of the fact that the SPA of men is 65, while it is substantially lower for most women in our sample and as low as 60 for women born before 1950. To illustrate simultaneous retirement planning, Figure 1 shows the distribution of the difference in the planned year of retirement between husband and wife. The left-hand panel illustrates the unconditional distribution and the right-hand panel conditions on the husband being at least 2 years older than his wife. The peak around zero indicates joint retirement planning, and the mass to the right of zero likely stems from men being older than women and women having a lower SPA. When conditioning on the husband being at least 2 years older than his wife in the right-hand panel, we see a substantial mass at 0 (same planned retirement year); we now also see a substantial mass at –2 (same planned retirement age).

Table 5 also shows that around 16% and 14% of men and women, respectively, are classified as highly skilled, and we see that men tend to visit the GP much less than women. Interestingly, however, men are more likely to expect their health to worsen in the future. The labor income of husbands is around £25,000, whereas that of the wives is on average around £14,000. Only around 13% of wives and 28% of husbands contribute to a private pension (PPP), while around 47% of wives and 51% of husbands are eligible to some occupational retirement scheme (EPP).

<sup>9</sup>After the relevant waves in the BHPS (11 and 16) were conducted, the Pension Act 2007 further increased the SPA for both men and women. Since the respondents were interviewed before this reform was passed (most interviews were done no later than 2006), we abstract from this and other subsequent reforms.



**FIGURE 1** Joint retirement planning. The figure illustrates the difference in the year of retirement between husband and wife. The peak around zero indicates joint retirement planning. Because the SPA of women is lower than that of men for most cohorts, it is expected that the distribution is right-tailed. The left-hand panel illustrates the unconditional distribution and the right-hand panel illustrates the distribution conditional on the husband being at least 2 years older than his spouse



**FIGURE 2** Model fit, individual retirement [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 4.2 | A model of retirement planning of dual-earner households

In this section, we formulate a discrete time version of the continuous time bivariate duration model proposed in Honoré and de Paula (2018). Specifically, we parametrize the difference in the utility flow between being retired and working. Utility maximization then gives an estimatable model for joint retirement planning of couples.

Consider first the husbands. We specify the difference in utility from being retired in period  $t$  compared to working as

$$U_h(t, t_w) = x'_h \beta_h + \delta_h(t) + \gamma \mathbf{1}_{\{C_h(t) \geq C_w(t_w)\}} + \varepsilon_h,$$

where  $C_h(t)$  is the calendar time,  $t_w$  is the retirement age of the wife, and  $C_w(t_w)$  thus is the calendar time at which the wife plans to retire. We interpret the term  $\gamma \mathbf{1}_{\{C_h(t) \geq C_w(t_w)\}}$  as a utility externality that allows the husband to enjoy a higher utility flow from planned retirement if the wife also plans to be retired at that time. We parametrize the planned retirement age function,  $\delta_h(t)$ , as a linear trend plus indicator functions for  $t \geq 55$ ,  $t \geq 60$  and  $t \geq 65$ . The histograms in Figure 2 suggest that these are empirically important. We interpret the first two as reflecting either social norms or heaping, while the third also reflects the fact that the SPA for men is 65.

Similarly, the difference in utility flow for the wife is

$$U_w(t, t_h) = x'_w \beta_w + \delta_w(t) + \gamma \mathbf{1}_{\{C_w(t) \geq C_h(t_h)\}} + \alpha \mathbf{1}_{\{t_w \geq SPA_w\}} + \varepsilon_w.$$

We again parametrize the function  $\delta_w(t)$  as a linear trend plus indicator functions for  $t \geq 55$ ,  $t \geq 60$  and  $t \geq 65$ . The term  $\alpha \mathbf{1}_{\{t \geq \text{SPA}_w\}}$  reflects the idea that for women there is variation in the SPA as discussed above. This allows one to infer the effect of the SPA separately from the dummies that reflect either heaping or institutional features (e.g., early and statutory retirement ages) at 55, 60, and 65.

To close the model, we assume that  $(\varepsilon_h, \varepsilon_w)$  is jointly normal with mean zero and covariance matrix  $\Omega$ , where the off-diagonal element of  $\Omega$  captures possibly correlated retirement preferences within households. We also assume that retirement is an absorbing state. When the difference in utility from retirement compared to working is increasing in age, this is not a binding constraint in the sense that individuals would not want to reenter the labor market once retired.

If a husband and a wife plan to retire at ages  $r_h$  and  $r_w$ , their discounted individual utilities are

$$V_h(r_h, r_w) = \sum_{t=r_h}^{T_{\max}} \rho^{t-\text{age}_h} U_h(t, r_w) \tag{Vh}$$

for a husband aged  $\text{age}_h$  and

$$V_w(r_w, r_h) = \sum_{t=r_w}^{T_{\max}} \rho^{t-\text{age}_w} U_w(t, r_h), \tag{Vw}$$

for a wife aged  $\text{age}_w$ . Finally, the optimal retirement plan for a household is determined jointly as

$$(R_h, R_w) = \arg \max_{r_h, r_w} \mathcal{A}[V_h(r_h, r_w), V_w(r_w, r_h)],$$

where  $\mathcal{A}(\cdot, \cdot)$  is a household aggregator. For the estimation, we choose  $\mathcal{A}(V_h, V_w) = V_h + \lambda V_w$  as in the Nash bargaining setting from Honoré and de Paula (2018) or, more generally, the collective model framework surveyed in Browning, Chiappori, and Weiss (2014).

It is clear that two-scale normalizations are necessary in order to estimate the model. First, the scale of  $\mathcal{A}$  cannot be identified and we therefore normalize the variance of  $\varepsilon_h$  to be  $\sigma_h^2 = 1$ . Secondly, the only effect of  $\lambda$  is to rescale all the parameters in  $V_w$ . We therefore normalize  $\lambda = 1$ . The model is thus in effect unitary.

Our parametrization is inspired by the ordered probit model. Consider the husbands. If  $\gamma = 0$  (such that there is no utility externality) and  $\delta_h$  is increasing, then the utility maximization will lead to planned retirement the first time  $x'_h \beta_h + \delta_h(t) + \varepsilon_h > 0$ . In other words, the chosen planned retirement age satisfies

$$-\delta_h(R_h) < x'_h \beta_h + \varepsilon_h \leq -\delta_h(R_h - 1),$$

which is exactly the ordered probit model. In that sense, the proposed model is a generalization of the ordered probit model to a bivariate case with simultaneity between the two outcomes.

### 4.3 | Indirect inference estimation

We estimate the model's parameter vector  $\theta = (\gamma, \alpha, \beta_h, \beta_w, \delta_h, \delta_w, \sigma_w^2, \sigma_{hw})$  through indirect inference<sup>10</sup>:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} g(\theta)' W g(\theta).$$

The weighting matrix,  $W$ , is diagonal with the inverse of the variances of the moments in the diagonal.  $g(\theta)$  is a  $K \times 1$  vector of differences between statistics/moments in the data and identical moments based on simulated data.

For each couple  $i$ , we simulate synthetic retirement plans by drawing  $S_{sim}$  vectors of taste shocks  $\varepsilon_i = \{\varepsilon_{i,h}^{(s)}, \varepsilon_{i,w}^{(s)}\}_{s=1}^{S_{sim}}$  from the joint normal distribution, and calculate the value of all combinations of retirement ages:

$$V_i^{(s)}(r_h, r_w) = V_h(r_h, r_w | x_i, \varepsilon_{i,h}^{(s)}, \varepsilon_{i,w}^{(s)}) + \lambda V_w(r_w, r_h | x_i, \varepsilon_{i,h}^{(s)}, \varepsilon_{i,w}^{(s)}),$$

<sup>10</sup>See, for example, Smith (1993), and Gouriéroux, Monfort, and Renault (1993). While we use the Wald criterion function, indirect inference can also be performed using other metrics (e.g., the likelihood ratio or Lagrange multiplier). See Smith (2008).

where the individual values are calculated as in Equations Vh and Vw. We then find the simulated retirement ages that maximize utility:

$$(R_{i,h}^{(s)}(\theta), R_{i,w}^{(s)}(\theta)) = \arg \max_{r_h, r_w} V_i^{(s)}(r_h, r_w)$$

for a given value of  $\theta$ .

To estimate the model parameters, we use four sets of auxiliary models/moments with a total of  $K = 52$  elements in  $g(\theta)$ . We describe in detail the construction of these moments in the Supporting Information and only list them here:

1. Ordinary least squares (OLS) coefficients from individual regressions of the planned retirement age on own and spousal covariates  $x_{i,h}$  and  $x_{i,w}$  together with indicators for the wife's birth cohort  $\mathbf{1}\{1950 < cohort_{w,i} \leq 1954\}$  and  $\mathbf{1}\{1955 \leq cohort_{w,i}\}$ .
2. The share of individuals planning to retire at ages 50–54, 55, 56–59, 60, 61–64, and 65, split by gender.
3. The covariance matrix of residuals from the regression in item 1 above for each household member.

		Husband		Wife	
$\gamma$	Joint leisure	0.026	(0.011)	0.026	(0.011)
$\alpha$	SPA age	—	—	0.105	(0.122)
<i>Explanatory variables (<math>\beta</math>)</i>					
	High skilled	-0.129	(0.100)	-0.148	(0.110)
	10+ GP visits	0.315	(0.291)	0.152	(0.157)
	Expect worse health	0.091	(0.112)	0.001	(0.109)
	Labor income (1,000£)	0.006	(0.003)	0.011	(0.005)
	Has private pension (PPP)	0.194	(0.092)	-0.005	(0.084)
	Has employer-provided pension (EPS)	0.610	(0.089)	-0.044	(0.060)
	Birth year (minus 1955)	0.005	(0.005)	-0.005	(0.007)
	Labor income (1,000£), spouse	0.005	(0.004)	0.003	(0.003)
	Has private pension (PPP), spouse	0.074	(0.093)	-0.005	(0.077)
	Has employer-provided pension (EPS), spouse	0.171	(0.076)	0.013	(0.080)
<i>Age variables (<math>\delta</math>)</i>					
	Constant	-2.413	(0.128)	-1.667	(0.474)
	Time trend (minus 25)	0.036	(0.004)	0.020	(0.007)
	Retirement age 55 dummy	0.632	(0.068)	0.729	(0.177)
	Retirement age 60 dummy	0.867	(0.038)	1.323	(0.362)
	Retirement age 65 dummy	1.978	(0.078)	1.452	(0.418)
$\sigma$	variance	1.000		0.917	
$\sigma_{hw}$	covariance	0.359		0.359	

TABLE 6 Estimation results, indirect inference

Note. The table reports the estimated simultaneous retirement planning model using the BHPS data and indirect inference. Asymptotic standard errors reported in parentheses.

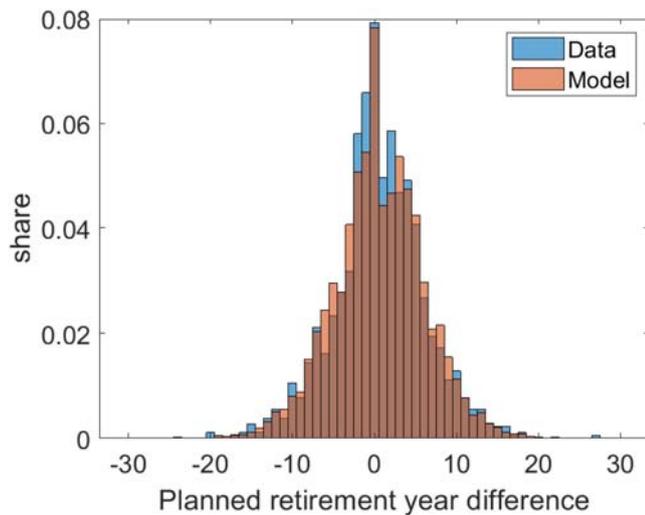


FIGURE 3 Model fit, joint retirement [Colour figure can be viewed at wileyonlinelibrary.com]

4. The share of couples with retirement plans such that (i) the wife plans to retire 1–2 years before her husband, (ii) the husband plans to retire 1–2 years before his wife, or (iii) the couple plan to retire in the same year.

The first set of moments is primarily included to help estimate  $\beta_h$ ,  $\beta_w$ , and  $\alpha$  in the utility function. The second set of moments is included primarily to help estimate the linear age trend and age dummies in  $\delta_h$  and  $\delta_w$ . The third set of moments is primarily included to estimate the covariance of the preference shocks for husband and wife,  $\Omega$ . Recall that we normalize  $\sigma_h^2 = 1$  and the remaining parameters in  $\Omega$  are thus  $\sigma_w^2$  and  $\sigma_{hw}$ . The final set of moments is included to estimate the value of joint leisure,  $\gamma$ . We will use our proposed sensitivity measures below to investigate these claims in a more systematic way.

#### 4.4 | Empirical results

We use the BHPS data discussed above to estimate the model of joint retirement planning of couples. We use the same moments as above and simulate  $S_{sim} = 2,000$  draws when approximating the expected moments. Table 6 reports the estimation results. We find a positive value of coordination of around  $\gamma \approx 0.026$ , around two to four times as large as the marginal utility from additional labor income of £1,000 and significant at the 5% level ( $p$ -value of 0.02).

Overall, the remaining statistically significant parameter estimates have the expected signs. High skilled individuals value retirement less. Less healthy people value retirement more, and having some form of pension savings increases the

TABLE 7 Sensitivity of  $\gamma$

Moment	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$
<i>Regression, husband</i>						
1 Constant	-0.006	0.259	0.000	-0.001	0.009	0.001
2 High skilled, husband	0.075	0.041	0.024	0.270	0.216	-0.010
3 10+ GP visits, husband	0.018	0.008	0.001	0.065	0.049	-0.002
4 Expect worse health, husband	-0.003	0.001	0.000	-0.008	0.009	0.000
5 Labor income, husband	0.007	0.000	0.000	-0.005	0.000	0.001
6 Has private pension, husband	-0.066	0.023	0.019	0.256	0.097	-0.004
7 Has employer provided pension, husband	-0.032	0.005	0.004	0.040	0.003	-0.005
8 Birth year (minus 1955), husband	-0.036	0.001	0.006	-0.006	0.001	0.004
9 High skilled, wife	0.015	0.001	0.001	-0.004	0.001	0.003
10 10+ GP visits, wife	-0.007	0.001	0.000	-0.002	0.001	0.002
11 Expect worse health, wife	-0.005	0.005	0.000	0.001	0.004	-0.001
12 Labor income, wife	-0.017	0.005	0.001	0.010	0.009	-0.001
13 Has private pension, wife	0.105	0.039	0.048	0.904	1.033	0.014
14 Has employer provided pension, wife	0.009	0.005	0.000	0.022	0.011	-0.001
15 Birth year, wife	0.003	0.003	0.000	0.001	0.000	-0.001
16 Birth year, wife in 1951–1955	-0.003	0.072	0.000	-0.002	0.030	0.002
17 Birth year, wife later than 1955	-0.005	0.130	0.000	-0.001	0.006	0.001
<i>Regression, wife</i>						
18 Constant	-0.024	0.075	0.002	-0.006	0.005	0.005
19 High skilled, husband	0.025	0.013	0.003	0.002	0.007	-0.000
20 10+ GP visits, husband	0.011	0.005	0.001	-0.002	0.004	0.002
21 Expect worse health, husband	-0.005	0.007	0.000	0.000	0.004	-0.000
22 Labor income, husband	-0.012	0.007	0.001	0.008	0.015	-0.001
23 Has private pension, husband	0.106	0.044	0.050	0.899	0.733	0.012
24 Has employer provided pension, husband	-0.062	0.012	0.017	0.003	0.048	0.009
25 Birth year (minus 1955), husband	0.023	0.090	0.002	0.009	0.032	-0.005
26 High skilled, wife	0.064	0.015	0.018	0.053	0.054	0.003
27 10+ GP visits, wife	0.009	0.001	0.000	0.033	0.010	-0.001
28 Expect worse health, wife	0.057	0.016	0.014	0.276	0.155	0.001
29 Labor income, wife	-0.017	0.007	0.001	0.028	0.025	-0.001
30 Has private pension, wife	-0.060	0.014	0.016	0.755	0.337	0.003
31 Has employer provided pension, wife	0.009	0.011	0.000	0.034	0.045	-0.001
32 Birth year, wife	-0.028	0.062	0.003	0.011	0.009	-0.005
33 Birth year, wife in 1951–1955	-0.013	0.039	0.001	0.003	0.013	-0.002
34 Birth year, wife later than 1955	-0.024	0.021	0.003	-0.003	0.001	0.002

(Continues)

TABLE 7 (Continued)

Moment		$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$
<i>Retirement age, husband</i>							
35	Share at ages 50–54	0.005	0.000	0.000	0.003	0.000	−0.001
36	Share at age 55	0.030	0.000	0.004	0.036	0.000	−0.009
37	Share at ages 56–59	−0.040	0.081	0.007	0.050	0.042	−0.009
38	Share at age 60	−0.003	0.043	0.000	−0.001	0.005	0.000
39	Share at ages 61–64	0.015	0.000	0.001	0.032	0.000	−0.003
40	Share at age 65	0.005	0.027	0.000	0.002	0.004	−0.000
<i>Retirement age, wife</i>							
41	Share at ages 50–54	0.024	0.011	0.002	0.007	0.007	−0.001
42	Share at age 55	0.010	0.067	0.000	0.007	0.019	−0.002
43	Share at ages 56–59	−0.024	0.014	0.003	−0.009	0.007	0.002
44	Share at age 60	−0.001	0.250	0.000	−0.001	0.020	0.000
45	Share at ages 61–64	0.024	0.040	0.003	0.058	0.037	−0.002
46	Share at age 65	0.006	0.082	0.000	0.004	0.018	−0.001
<i>Simultaneous retirement</i>							
47	var( $e_h$ )	−0.008	0.002	0.000	−0.004	0.000	0.001
48	var( $e_w$ )	−0.005	0.076	0.000	−0.005	0.021	0.001
49	cov( $e_h, e_w$ )	−0.145	0.204	0.092	0.757	0.557	−0.040
50	diff [−2,−1]	0.018	0.035	0.001	0.018	0.037	−0.007
51	diff [1,2]	−0.113	0.000	0.056	−0.067	0.000	0.077
52	Joint retirement	0.343	0.684	0.516	8.019	5.541	−0.030

Note. The table reports the sensitivity measures of  $\gamma$  for the estimated joint retirement planning model.

value of retirement. Having an employer-provided pension (EPS) especially increases the utility from retirement compared to working for husbands. Perhaps surprisingly, we find that higher earning women value retirement more but this could proxy for higher wealth, which could lead to a higher propensity to retire. All spousal variables seem to matter less and are not statistically significant at most common significance levels. Interestingly, we estimate a small positive and insignificant increase in the expected retirement age of women in response to an increased SPA. This goes in line with other studies finding a relatively low degree of awareness of the reform (Crawford & Tetlow, 2010).

Figure 2 shows the histogram of planned retirement ages for women and men. We see that the model does a quite good job fitting the empirical distribution. Likewise, Figure 3 shows the empirical and predicted distribution of retirement year differences between couples. The predicted distribution matches the empirical one well, although there are small deviations.

Table 7 shows the proposed sensitivity measures together with the one proposed by Andrews et al. (2017). We only report the measures for the parameter of interest here: the value of joint leisure,  $\gamma$ . All reported measures are scaled as discussed in Section 2. The measure proposed by Andrews et al. (2017) is scaled such that  $\mathcal{E}_1^{(j,k)} = M_1^{(j,k)} \sqrt{S^{(k,k)}}$ .

Clearly, the moments which  $\gamma$  is most sensitive to are related to simultaneous retirement. In particular, we see from  $\mathcal{E}_4$  and  $\mathcal{E}_5$  that leaving out the moment “the share planning to retire the same year” (moment 52) when estimating the model would increase the asymptotic variance of  $\gamma$  by a factor of 8. This confirms the intuition that this moment is extremely informative about the value of joint leisure. The share retiring within 2 years difference also seems important. In particular, the correlation between the OLS regression residuals are important. This is also intuitive since this moment captures a combination of correlated shocks and preferences for joint leisure.

## 5 | CONCLUDING REMARKS

Structural econometric models are often estimated by matching moments that depend on the parameters and on the data in a highly nonlinear way. This can make it difficult to develop intuition for which moments of the data are informative about which parameter. In this paper, we have proposed a number of very simple sensitivity measures that are meant to shed light on this.

We have illustrated our measures in two artificial examples. The first is a simple probit model and the second a mixed proportional hazard model with time-varying covariates. The first illustrates that the proposed measures are reasonable in a setting where the answer is rather obvious ex ante. The second is chosen because it illustrates how the measures can be used to gain insights, which are not so obvious.

We also illustrated the measures in a simple structural econometric model of household retirement planning. This application is of independent interest because it highlights the importance of modeling wives' and husbands' retirement decisions jointly.

The econometric model for retirement that we develop can be interpreted as a bivariate ordered-choice model with simultaneity. Specifically, if the “utility externality” parameter is 0, then the model that we estimate simplifies to a bivariate ordered probit model. This may make it tractable in other applications.

## ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the National Science Foundation (grant numbers SES-1530741 and SES-1824131), the Danish Council for Independent Research in Social Sciences (FSE, grant number 4091-00040), the ESRC through the Centre for Microdata Methods and Practice (RES-589-28-0001) and the ERC (SG338187). The activities of the Center for Economic Behavior and Inequality (CEBI) are financed by a grant from the Danish National Research Foundation. We thank Sharada Dharmasankar, the co-editor and three anonymous referees for constructive comments and suggestions. A previous version of this paper was distributed under the title “Sensitivity of Estimation Precision to Moments with an Application to a Model of Joint Retirement Planning of Couples.”

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [<http://qed.econ.queensu.ca/jae/datasets/honore001/>].

## REFERENCES

- Altonji, J. G., & Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14(3), 353–366.
- An, M. Y., Christensen, B. J., & Gupta, N. D. (2004). Multivariate mixed proportional hazard modelling of the joint retirement of married couples. *Journal of Applied Econometrics*, 19(6), 687–704.
- Andrews, I., Gentzkow, M., & Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics*, 132(4), 1553–1592.
- Armstrong, T. B., & Kolesár, M. (2018). Sensitivity analysis using approximate moment condition models (Working paper). arXiv:1808.07387.
- Banks, J., Blundell, R., & Casanova, M. (2010). The dynamics of retirement behavior in couples: Reduced-form evidence from England and the US (Working paper). Los Angeles, CA: Department of Economics, UCLA.
- Blau, D. M. (1998). Labor force dynamics of older married couples. *Journal of Labor Economics*, 16(3), 595–629.
- Blau, D. M., & Gilleskie, D. B. (2006). Health insurance and retirement of married couples. *Journal of Applied Econometrics*, 21(7), 935–953.
- Blundell, R., Meghir, C., & Smith, S. (2004). Pension incentives and the pattern of retirement in the United Kingdom. In J. Gruber, & D. A. Wise (Eds.), *Social security programs and retirement around the world: Micro-estimation* (pp. 643–689). Chicago, IL: University of Chicago Press.
- Bonhomme, S., & Weidner, M. (2018). Minimizing sensitivity to model misspecification (Working paper). arXiv:1807.02161.
- Bozio, A., Crawford, R., & Tetlow, G. (2010). The history of state pensions in the UK: 1948 to 2010 (IFS Briefing Note BN15). London, UK: Institute for Fiscal Studies.
- Browning, M., Chiappori, P., & Weiss, Y. (2014). *Economics of the family*. Cambridge, UK: Cambridge University Press.
- Casanova, M. (2010). Happy together: A structural model of couples' joint retirement choices (Working paper). Los Angeles, CA: Department of Economics, UCLA.
- Christensen, T., & Connault, B. (2019). Counterfactual sensitivity and robustness (Working paper). arXiv:1904.00989.
- Coile, C. (2004). Retirement incentives and couples' retirement decisions. *Topics in Economic Analysis and Policy*, 4(1), article 17.
- Crawford, R., & Tetlow, G. (2010). Employment, retirement and pensions. In J. Banks, C. Lessof, J. Nazroo, N. Rogers, M. Stafford, & A. Steptoe (Eds.), *Financial circumstances, health and well-being of the older population in England: The 2008 English longitudinal study of ageing (wave 4)* (pp. 11–75). London, UK: Institute for Fiscal Studies.
- Cribb, J., Emmerson, C., & Tetlow, G. (2013). Incentives, shocks or signals: Labour supply effects of increasing the female state pension age in the UK (Working Paper No. 13/03). London, UK: Institute for Fiscal Studies.
- de Grip, A., Fouarge, D., & Montizaan, R. (2013). How sensitive are individual retirement expectations to raising the retirement age? *De Economist*, 161, 225–251.
- Eisenhauer, P., Heckman, J. J., & Mosso, S. (2015). Estimation of dynamic discrete choice models by maximum likelihood and the simulated method of moments. *International Economic Review*, 56(2), 331–357.
- Gayle, G. L., & Shephard, A. (2019). Optimal taxation, marriage, home production, and family labor supply. *Econometrica*, 87(1), 291–326.
- Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8, 85–118.

- Gustman, A. L., & Steinmeier, T. L. (2000). Retirement in dual-career families: A structural model. *Journal of Labor Economics*, 18(3), 503–545.
- Gustman, A. L., & Steinmeier, T. L. (2004). Social security, pensions and retirement behaviour within the family. *Journal of Applied Econometrics*, 19(6), 723–737.
- Hahn, J. (1994). The efficiency bound of the mixed proportional hazard model. *Review of Economic Studies*, 61(4), 607–629.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054.
- Honoré, B. (1990). Simple estimation of a duration model with unobserved heterogeneity. *Econometrica*, 58(2), 453–473.
- Honoré, B. E., & de Paula, A. (2018). A new model for interdependent durations. *Quantitative Economics*, 9(3), 1299–1333.
- Hurd, M. D. (1990). The joint retirement decision of husbands and wives. In D. A. Wise (Ed.), *Issues in the economics of aging* (pp. 231–258). Chicago, IL: University of Chicago Press.
- Jia, Z. (2005). Labor supply of retiring couples and heterogeneity in household decision-making structure. *Review of Economics of the Household*, 3, 215–233.
- Moen, P., Huang, Q., Plassmann, V., & Dentinger, E. (2006). Deciding the future: Do dual-earner couples plan together for retirement?. *American Behavioral Scientist*, 49(10), 1422–1443.
- Pienta, A. M., & Hayward, M. D. (2002). Who expects to continue working after age 62? The retirement plans of couples. *Journal of Gerontology: Social Sciences*, 57B(4), 199–208.
- Smith, A. (1993). Estimating nonlinear time series models using vector-autoregressions: Two approaches. *Journal of Applied Econometrics*, 8, S63–S84.
- Smith, A. (2008). Indirect inference. In *The new Palgrave dictionary of economics*. London, UK: Palgrave Macmillan.
- Thurley, D., & Keen, R. (2017). State pension age increases for women born in the 1950s (Commons Briefing papers No. CBP-7405). London, UK: House of Commons Library.
- van der Klaauw, W., & Wolpin, K. I. (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1–2), 21–42.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Honoré B, Jørgensen T, de Paula Á. The informativeness of estimation moments. *J Appl Econ*. 2020;1–17. <https://doi.org/10.1002/jae.2779>