The Informativeness of Estimation Moments*

Bo E. Honoré[†] Thomas H. Jørgensen[‡] Áureo de Paula[§]

April 24, 2020

Abstract

This paper introduces measures for how each moment contributes to the precision of parameter estimates in GMM 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 UK-BHPS, 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.

Keywords: Parameter Sensitivity, GMM, Ordered Models, Retirement, Externalities.
JEL Code: C01, C13, C35, J32.

---

*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 no. 4091-00040), the ESRC through the Centre for Microdata Methods and Practice (RES-589-28-0001) and the ERC (SG338187). The activities of Center for Economic Behavior and Inequality (CEBI) are financed by a grant from the Danish National Research Foundation. We thank Sharada Dharmasankar, the coeditor 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."

[†]Department of Economics, Princeton University and The Dale T. Mortensen Centre at the University of Aarhus. E-mail: honore@princeton.edu

[‡]Center for Economic Behavior and Inequality (CEBI), Department of Economics, University of Copenhagen. E-mail: thomas.h.jorgensen@econ.ku.dk. Webpage: www.tjeconomics.com.

[§]Department of Economics, University College London, C*e*MMAP and IFS, London, UK. E-mail: apaula@ucl.ac.uk

1 Introduction

Indirect inference and other nonlinear 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) have also studied local misspecification. Christensen and Connault (2019) studied global misspecification.

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,[1] the empirical evidence of joint retirement *planning* of couples is much more scarce and with ambiguous findings.[2] 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 $E\left[f\left(x_i,\theta_0\right)\right]=0$, where x_i is data for observation i and it is assumed that this defines a unique θ_0. The generalized method of moments (GMM) estimator of θ_0 is $\widehat{\theta}=\arg\min_\theta\left(\frac{1}{n}\sum_{i=1}^n f\left(x_i,\theta\right)\right)' W_n\left(\frac{1}{n}\sum_{i=1}^n f\left(x_i,\theta\right)\right)$, 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 over-identified models where the number of moments are larger than the number of parameters in θ, and the weighting matrix thus plays a role.

[1] See e.g. Hurd (1990); Blau (1998); Gustman and Steinmeier (2000); Gustman and Steinmeier (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 and de Paula (2018).

[2] See Pienta and Hayward (2002); Moen, Huang, Plassmann and Dentinger (2006); and de Grip, Fouarge and Montizaan (2013).

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

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

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

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

where

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

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

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

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.

Andrews, Gentzkow and Shapiro (2017) proposed the sensitivity measure

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

It is clear from (1) that M_1 provides the mapping from moment misspecification of the type $E\left[f\left(x_i, \theta_0\right)\right] = \rho \neq 0$ into parameter biases for small ρ. Alternatively, by noting that $\Sigma = M_1 S M_1'$, M_1 tells us how additional noise in each of the sample moments $\frac{1}{n}\sum_{i=1}^{n} f\left(x_i, \theta_0\right)$ would result in additional noise in each element of $\widehat{\theta}$. 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, Gentzkow and Shapiro (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 $\widehat{\theta}$ 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

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

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 non-optimal) weighting matrix unchanged. This measure is

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

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

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

Here $\odot$ denotes element-wise multiplication and ι_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 over-identified 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 $(\widehat{G}'\tilde{W}_k\widehat{G})$ 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 over-identified.

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 and 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 reparameterization of those reduced form models would lead to moment conditions which 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 kth moment? This measure is formally defined as the derivative

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

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 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

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

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 non-optimal weighting matrix because it is very common in empirical applications.[3]

[3] There are many examples of this. This includes 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.,

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. ε_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 $P(y_i = 1) = 0.66$.[4]

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:

$$\begin{pmatrix} E[e(\theta)] & E[e(\theta)\, x_1] & E[e(\theta)\, x_2] & E[e(\theta)\, x_1^2] & E[e(\theta)\, x_1 x_2] & E[e(\theta)\, x_2^2] \end{pmatrix}'$$

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 θ_0. Moreover, we expect the first moment to be the most important for determining $\widehat{\beta}_0$, and the second and third for determining $\widehat{\beta}_1$ and $\widehat{\beta}_2$, 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.[5] We think of the latter as a practical alternative to the efficient weighting matrix.

It is clear from Table 1 that the first three moments are indeed the most informative about

$W = I$) 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.

[4] We also supply Python code to illustrate our approach.

[5] We illustrate the proposed sensitivity measures through Monte Carlo simulation of the expected values using 10^7 simulated observations.

β_0, β_1 and β_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 the elements 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 θ_0 is driven by the first three moments.[6] Adding noise to the last three moments has essentially no effect on the precision of the optimal GMM estimator of θ_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 non-optimal 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 β_1 by around 400 percent. This confirms that $E[ex_1]$ is instrumental for precise estimation of β_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 θ, 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, which follows a mixed proportional hazard model with time-varying

[6] $\mathcal{E}_2$ in Tables 1 and 2 differ only because of simulation error.

covariates and a Weibull as the baseline hazard

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

where α is the scale parameter which 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, η captures unobserved heterogeneity. Except for moment assumptions, no assumptions are made on the distribution of η.

We then have the survival function for T,

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

Since

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

we have

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

or

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

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

Equation (2) suggests moment conditions of the type

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

for functions of the covariates, ψ. Here, β_0 captures the mean of $-\log(\eta)$ which is assumed to be finite.

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

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

or

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

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

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

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, η, follows a log-normal distribution, where the underlying normal has mean 0 and variance 1/2. η is independent of $x(\cdot)$. Finally, $\beta=\left(-1,1/\sqrt{2},1/\sqrt{2}\right)'$ and $\alpha=2$. With this, the median duration is approximately 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

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

where we use the five moments given by (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 non-zero with respect to θ_0 and θ_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 α, 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 α. 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.

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 through 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.[7] 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).[8] Finally, we define individuals as highly skilled if they have completed the first or second stage of tertiary education (ISCED codes 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 to 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.

The State Pension Age (SPA)

The state pension age (SPA) in the U.K. is the age where individuals become eligible to receive state pension from the government. Individuals who have reached SPA and contributed to the scheme for sufficiently many years are eligible to receive a weekly 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 U.K.

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,

[7]The exact formulation in wave 11 is slightly different: *"At what age do you expect to retire/will you consider yourself to be retired?"*

[8]The EPP includes both defined and contributed benefit (DB and CB) plans and we cannot distinguish between them. Blundell, Meghir and Smith (2004) show, however, that DB plans were most common in the U.K. in this period.

1950. While the SPA for men was unaffected, the SPA for women was gradually increased by one 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.[9] Since this might affect individual expectations, our modelling framework explicitly allows for an effect of the Pension Act 1995 on retirement planning.

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 percent of our sample. Husbands in the estimation sample are approximately 1.5 years older than their wives, plan to retire two 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 panel illustrates the unconditional distribution and the right 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 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 percent 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 $£25,000$ while that of the wives is on average around $£14,000$. Only around 13 percent of wives and 28 percent of husbands contribute to a private pension (PPP), while around 47 percent of wives and 51 percent of husbands are eligible to some

[9] 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 was done no later than 2006), we abstract from this and other subsequent reforms.

occupational retirement scheme (EPP).

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 parameterize 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

$$U_h(t, t_w) = x_h'\beta_h + \delta_h(t) + \gamma \mathbf{1}_{\{\mathcal{C}_h(t) \geq \mathcal{C}_w(t_w)\}} + \varepsilon_h,$$

where $\mathcal{C}_h(t)$ is the calendar time, t_w is the retirement age of the wife, and $\mathcal{C}_w(t_w)$ thus is the calendar time at which the wife plans to retire. We interpret the term $\gamma \mathbf{1}_{\{\mathcal{C}(t) \geq \mathcal{C}(t_w)\}}$ 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 parameterize 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 below 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

$$U_w(t, t_h) = x_w'\beta_w + \delta_w(t) + \gamma \mathbf{1}_{\{\mathcal{C}_w(t) \geq \mathcal{C}_h(t_h)\}} + \alpha \mathbf{1}_{\{t_w \geq SPA_w\}} + \varepsilon_w.$$

We again parameterize 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 $\alpha \mathbf{1}_{\{t_w \geq SPA_w\}}$ 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 Ω, where the off-diagonal element of Ω 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 re-enter 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

$$V_h(r_h, r_w) = \sum_{t=r_h}^{T_{max}} \rho^{t-age_h} U_h(t, r_w)$$

for a husband aged age_h and

$$V_w(r_w, r_h) = \sum_{t=r_w}^{T_{max}} \rho^{t-age_w} U_w(t, r_h),$$

for a wife aged 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 ε_h to be $\sigma_h^2 = 1$. Secondly, the only effect of λ is to re-scale all the parameters in V_w. We therefore normalize $\lambda = 1$. The model is thus in effect unitary.

Our parameterization is inspired by the ordered probit model. Consider the husbands. If $\gamma = 0$ (such that there is no utility externality) and δ_h is increasing, then the utility maximation 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[10],

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

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

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

where the individual values are calculated as in (**??**) and (**??**). 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 θ.

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 supplemental material and only list them here:

1. 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.

[10] 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 (for example, the likelihood ratio or Lagrange multiplier). See Smith (2008).

3. The covariance matrix of residuals from the regression in bullet 1 above for each household member.

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 are primarily included to help estimate β_h, β_w, and α in the utility function. The second set of moments are included primarily to help estimate the linear age trend and age dummies in δ_h and δ_w. The third set of moments are primarily included to estimate the covariance of the preference shocks for husband and wife, Ω. Recall that we normalize $\sigma_h^2 = 1$ and the remaining parameters in Ω are thus σ_w^2 and σ_{hw}. The final set of moments are included to estimate the value of joint leisure, γ. 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 $S_{sim} = 2000$ 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 increase the 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 and 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 show the proposed sensitivity measures together with the one proposed by Andrews, Gentzkow and Shapiro (2017). We only report the measures for the parameter of interest here: The value of joint leisure, γ. All reported measures are scaled as discussed in Section 2. The measure proposed by Andrews, Gentzkow and Shapiro (2017) is scaled such that $\mathcal{E}_1^{(j,k)} = M_1^{(j,k)}\sqrt{S^{(k,k)}}$.

Clearly, the moments which γ 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 γ 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 modelling 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.

References

Altonji, J. G. and L. M. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14(3), 353–366.

An, M. Y., B. J. Christensen and N. D. Gupta (2004): "Multivariate mixed proportional hazard modelling of the joint retirement of married couples," *Journal of Applied Econometrics*, 19(6), 687–704.

Andrews, I., M. Gentzkow and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132(4), 1553–1592.

Armstrong, T. B. and M. Kolesár (2018): "Sensitivity Analysis using Approximate Moment Condition Models," Working paper, arXiv:1808.07387.

Banks, J., R. Blundell and M. Casanova (2010): "The dynamics of retirement behavior in couples: Reduced-form evidence from England and the US," Working paper, 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. and D. B. Gilleskie (2006): "Health Insurance and Retirement of Married Couples," *Journal of Applied Econometrics*, 21(7), 935–953.

Blundell, R., C. Meghir and S. Smith (2004): "Pension Incentives and the Pattern of Retirement in the United Kingdom," in *Social Security Programs and Retirement around the*

World: Micro-Estimation, ed. by J. Gruber and D. A. Wise, chap. 11, pp. 643–689. University of Chicago Press.

Bonhomme, S. and M. Weidner (2018): "Minimizing Sensitivity to Model Misspecification," Working paper, arXiv:1807.02161.

Bozio, A., R. Crawford and G. Tetlow (2010): "The history of state pensions in the UK: 1948 to 2010," IFS Briefing Note BN105, Institute for Fiscal Studies.

Browning, M., P. Chiappori and Y. Weiss (2014): *Economics of the Family*. Cambridge University Press.

Casanova, M. (2010): "Happy Together: A Structural Model of Couples' Joint Retirement Choices," Working paper, Department of Economics, UCLA.

Christensen, T. and B. Connault (2019): "Counterfactual Sensitivity and Robustness," Working paper, arXiv:1904.00989.

Coile, C. (2004): "Retirement Incentives and Couples' Retirement Decisions," *Topics in Economic Analysis & Policy*, 4(1), article 17.

Crawford, R. and G. Tetlow (2010): "Employment, retirement and pensions," in *Financial circumstances, health and well-being of the older population in England: The 2008 English Longitudinal Study of Ageing (Wave 4)*, ed. by J. Banks, C. Lessof, J. Nazroo, N. Rogers, M. Stafford and A. Steptoe. Institute for Fiscal Studies.

Cribb, J., C. Emmerson and G. Tetlow (2013): "Incentives, shocks or signals: labour supply effects of increasing the female state pension age in the UK," Working Paper 13/03, Institute for Fiscal Studies.

de Grip, A., D. Fouarge and R. Montizaan (2013): "How Sensitive are Individual Retirement Expectations to Raising the Retirement Age?," *De Economist*, 161, 225–251.

Eisenhauer, P., J. J. Heckman and S. Mosso (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. and A. Shephard (2019): "Optimal Taxation, Marriage, Home Production, and Family Labor Supply," *Econometrica*, 87(1), 291–326.

Gouriéroux, C., A. Monfort and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

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

——— (2004): "Social Security, Pensions and Retirement Behaviour 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).

Honoré, B. E. and Â. de Paula (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 *Issues in the Economics of Aging*, ed. by D. A. Wise, chap. 8, pp. 231–258. 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., Q. Huang, V. Plassmann and E. Dentinger (2006): "Deciding the Future: Do Dual-Earner Couples Plan Together for Retirement?," *American Behavioral Scientist*, 49(10), 1422–1443.

Pienta, A. M. and M. D. Hayward (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.

——— (2008): *Indirect Inference*. The New Palgrave Dictionary of Economics.

Thurley, D. and R. Keen (2017): "State Pension age increases for women born in the 1950s," Commons Briefing papers CBP-7405, House of Commons Library.

van der Klaauw, W. and K. I. Wolpin (2008): "Social security and the retirement and savings behavior of low-income households," *Journal of Econometrics*, 145(1-2), 21–42.

Table 1: Sensitivity Measures, Probit Model, Optimal Weighting

	Moment					
	$\mathbb{E}[e]$	$\mathbb{E}[ex_1]$	$\mathbb{E}[ex_2]$	$\mathbb{E}\left[ex_1^2\right]$	$\mathbb{E}[ex_1x_2]$	$\mathbb{E}\left[ex_2^2\right]$
			M_1			
β_0	4.261	1.475	1.469	0.192	0.378	0.184
β_1	1.190	6.570	−1.286	0.223	0.141	−0.069
β_1	1.193	−1.286	6.567	−0.073	0.152	0.214
			$\mathcal{E}_2$			
β_0	1.104	0.088	0.087	0.003	0.004	0.003
β_1	0.060	1.207	0.046	0.003	0.000	0.000
β_1	0.060	0.046	1.205	0.000	0.000	0.003
			$\mathcal{E}_3$			
β_0	1.104	0.088	0.087	0.003	0.004	0.003
β_1	0.060	1.207	0.046	0.003	0.000	0.000
β_1	0.060	0.046	1.205	0.000	0.000	0.003
			$\mathcal{E}_4$			
β_0	1.206	0.292	0.291	0.005	0.003	0.005
β_1	0.065	4.014	0.155	0.004	0.006	0.000
β_1	0.065	0.153	4.034	0.000	0.006	0.004
			$\mathcal{E}_5$			
β_0	1.203	0.292	0.291	0.001	0.003	0.001
β_1	0.065	4.014	0.155	0.001	0.000	0.000
β_1	0.065	0.153	4.034	0.000	0.000	0.001
			$\mathcal{E}_6$			
β_0	0.000	0.000	0.000	0.000	0.000	0.000
β_1	0.000	0.000	0.000	0.000	0.000	0.000
β_1	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Simulations based on 10^7 *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			
β_0	3.374	1.633	1.630	1.036	−0.681	1.035
β_1	1.354	5.656	−1.185	−0.853	−1.360	0.882
β_1	1.351	−1.185	5.658	0.881	−1.360	−0.851
			$\mathcal{E}_2$			
β_0	1.104	0.088	0.087	0.003	0.004	0.003
β_1	0.060	1.207	0.046	0.003	0.000	0.000
β_1	0.060	0.046	1.205	0.000	0.000	0.003
			$\mathcal{E}_3$			
β_0	0.651	0.101	0.101	0.080	0.013	0.080
β_1	0.071	0.817	0.036	0.037	0.034	0.039
β_1	0.070	0.036	0.817	0.039	0.034	0.037
			$\mathcal{E}_4$			
β_0	1.076	0.341	0.340	−0.010	−0.011	−0.011
β_1	0.042	3.783	0.116	−0.038	−0.031	−0.028
β_1	0.042	0.114	3.802	−0.028	−0.032	−0.038
			$\mathcal{E}_5$			
β_0	1.203	0.292	0.291	0.001	0.003	0.001
β_1	0.065	4.014	0.155	0.001	0.000	0.000
β_1	0.065	0.153	4.034	0.000	0.000	0.001
			$\mathcal{E}_6$			
β_0	−0.101	0.002	0.002	0.040	0.017	0.041
β_1	0.011	−0.142	0.002	0.044	0.048	0.037
β_1	0.011	0.002	−0.142	0.037	0.048	0.044

Notes: Simulations based on 10^7 *observations.*

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		
β_0	−0.503	0.001	3.375	−2.053	−1.934
β_1	−0.000	−0.500	0.018	−0.015	−0.014
β_2	−0.000	0.000	−0.228	−0.251	−0.184
α	−0.019	0.009	24.478	−15.092	−14.181
			$\mathcal{E}_2$		
β_0	0.028	0.000	1.282	0.474	0.421
β_1	0.000	0.998	0.001	0.001	0.001
β_2	0.000	0.000	0.155	0.187	0.100
α	0.000	0.000	1.299	0.494	0.436
			$\mathcal{E}_3$		
β_0	0.028	0.000	1.282	0.474	0.421
β_1	0.000	0.998	0.001	0.001	0.001
β_2	0.000	0.000	0.155	0.187	0.100
α	0.000	0.000	1.299	0.494	0.436
			$\mathcal{E}_4$		
β_0	>100*	>100*	4.841	0.196	0.274
β_1	0.324*	>100*	0.005	0.000	0.001
β_2	0.012*	>100*	0.584	0.077	0.065
α	3.935*	>100*	4.904	0.203	0.284
			$\mathcal{E}_5$		
β_0	>100*	>100*	4.841	0.196	0.274
β_1	0.324*	>100*	0.005	0.000	0.001
β_2	0.012*	>100*	0.584	0.077	0.065
α	3.935*	>100*	4.904	0.203	0.284
			$\mathcal{E}_6$		
β_0	0.000	−0.000	0.000	0.000	0.000
β_1	−0.000	−0.000	0.000	0.000	0.000
β_2	0.000	0.000	−0.000	0.000	0.000
α	−0.000	−0.000	0.000	0.000	0.000

Notes: Simulations based on 10^7 *observations.*

* 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.

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		
β_0	−0.503	0.001	3.066	−1.117	−2.679
β_1	−0.000	−0.500	0.016	−0.010	−0.019
β_2	−0.000	0.000	−0.234	−0.234	−0.197
α	−0.021	0.009	22.219	−8.255	−19.619
			$\mathcal{E}_2$		
β_0	0.028	0.000	1.282	0.474	0.421
β_1	0.000	0.998	0.001	0.001	0.001
β_2	0.000	0.000	0.155	0.187	0.100
α	0.000	0.000	1.299	0.494	0.436
			$\mathcal{E}_3$		
β_0	0.027	0.000	1.017	0.135	0.775
β_1	0.000	0.998	0.001	0.000	0.001
β_2	0.000	0.000	0.162	0.163	0.115
α	0.000	0.000	1.027	0.142	0.800
			$\mathcal{E}_4$		
β_0	>100*	>100*	4.612	0.149	0.224
β_1	0.323*	>100*	0.005	0.000	0.000
β_2	0.011*	>100*	0.583	0.077	0.065
α	3.737*	>100*	4.667	0.155	0.232
			$\mathcal{E}_5$		
β_0	>100*	>100*	4.841	0.196	0.274
β_1	0.324*	>100*	0.005	0.000	0.001
β_2	0.012*	>100*	0.584	0.077	0.065
α	3.935*	>100*	4.904	0.203	0.284
			$\mathcal{E}_6$		
β_0	0.000	−0.000	−0.049	−0.054	0.102
β_1	0.000	0.000	−0.000	−0.000	0.000
β_2	0.000	−0.000	0.002	−0.006	0.004
α	0.000	−0.000	−0.049	−0.056	0.105

Notes: Simulations based on 10^7 *observations.*

* 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.

Table 5: Descriptive Statistics

	Mean	Std.	Min	Max	Obs.
Age, husband	49.613	5.53	40	59	1730
Age, wife	48.128	5.34	40	59	1730
Planned retirement age, husband	62.606	3.87	50	70	1730
Planned retirement age, wife	60.301	3.72	50	70	1730
Diff. in planned retirement year (husband-wife)	0.823	5.71	-20	27	1730
High skilled, husband	0.157	0.36	0	1	1730
High skilled, wife	0.139	0.35	0	1	1730
10+ GP visits, husband	0.039	0.19	0	1	1729
10+ GP visits, wife	0.080	0.27	0	1	1729
Expect worse health, husband	0.182	0.39	0	1	1641
Expect worse health, wife	0.115	0.32	0	1	1645
Labor income (£1,000), husband	25.248	17.12	0	244	1600
Labor income (£1,000), wife	13.815	10.78	0	109	1442
Private pension, husband	0.280	0.45	0	1	1730
Private pension, wife	0.134	0.34	0	1	1730
Employer pension, husband	0.514	0.50	0	1	1730
Employer pension, wife	0.466	0.50	0	1	1730

Table 6: Estimation Results, Indirect Inference

		Husband		Wife	
γ	Joint leisure	0.026	(0.011)	0.026	(0.011)
α	SPA age	–	–	0.105	(0.122)
Explanatory variables (β)					
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)
Age variables (δ)					
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)
σ	variance	1.000		0.917	
σ_{hw}	covariance	0.359		0.359	

Notes: The table reports the estimated simultaneous retirement planning model using the BHPS data using indirect inference. Asymptotic standard errors reported in brackets.

Table 7: Sensitivity of γ

	Moment	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$
Regression, husband							
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
Regression, wife							
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

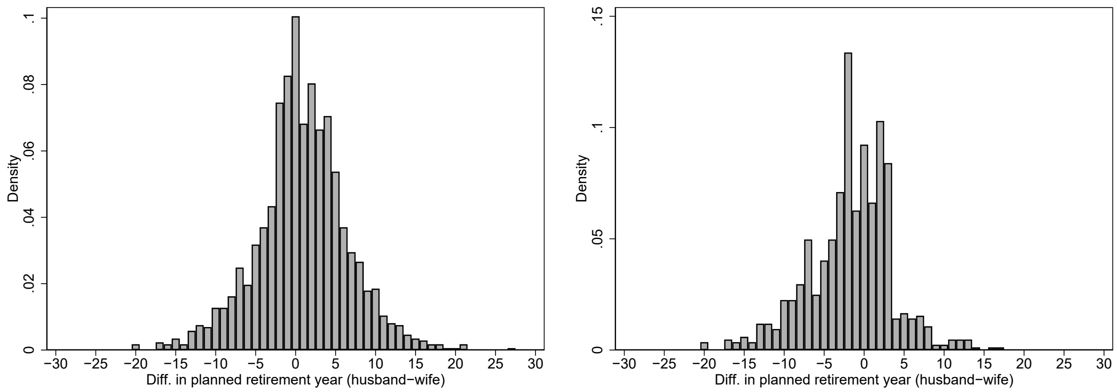
Notes: The table reports the sensitivity measures of γ for the estimated joint retirement planning model.

Table 7: Sensitivity of γ (continued)

Moment		$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$
Retirement age, husband							
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
Retirement age, wife							
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
Simultaneous retirement							
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

Notes: The table reports the sensitivity measures of γ for the estimated joint retirement planning model.

Figure 1: Joint Retirement Planning



Notes: Figure 1 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 from that of men for most cohorts, it is expected that the distribution is right-tailed. The left panel illustrates the unconditional distribution and the right panel illustrates the distribution conditional on the husband being at least 2 years older than his spouse.

Figure 2: Model Fit, Individual Retirement

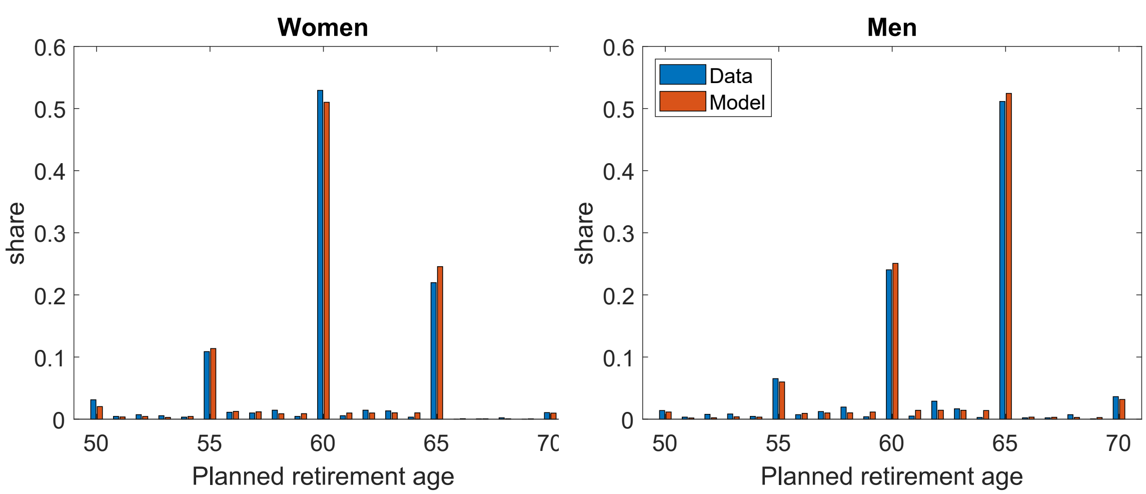
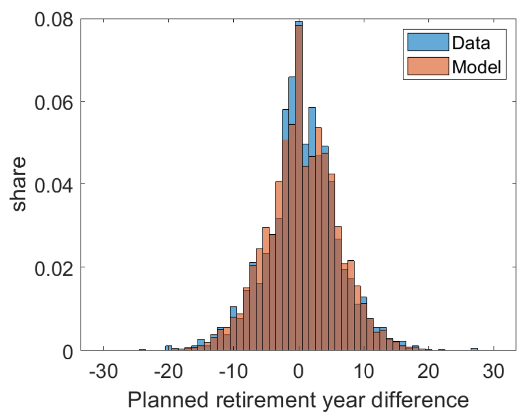


Figure 3: Model Fit, Joint Retirement



Online supplemental material

Definition of Moments used for Estimation

Individual OLS Moment Conditions. Let $R_{i,j}$ denote the planned retirement age of member j in household i and $X_i = (1, x'_{i,h}, x'_{i,w}, \mathbf{1}\{1950 < cohort_{w,i} \leq 1954\}, \mathbf{1}\{1955 \leq cohort_{w,i}\})'$ denote the set of control variables. We include as the first set of moments

$$\mathcal{M}_1(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\begin{pmatrix} X_i e_{i,h}^{(s)}(\theta) \\ X_i e_{i,w}^{(s)}(\theta) \end{pmatrix}$$

where, for $j = \{h, w\}$,

$$e_{i,j}^{(s)}(\theta) = R_{i,j}^{(s)}(\theta) - X_i'\hat{\beta}_j^{OLS}$$

where $\hat{\beta}_j^{OLS} = (X'X)^{-1}X'R_j$ are the OLS regression coefficients using the data.

Covariance Matrix of Regression Residuals. The second set of moments are related to the regression above. Particularly, we include as the second set of moments the simulated difference in the moments of the error terms

$$\mathcal{M}_2(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\begin{pmatrix} e_{i,h}^2 - (e_{i,h}^{(s)}(\theta))^2 \\ e_{i,w}^2 - (e_{i,w}^{(s)}(\theta))^2 \\ e_{i,h}e_{i,w} - e_{i,h}^{(s)}(\theta)e_{i,w}^{(s)}(\theta) \end{pmatrix}$$

where $e_{i,j} = R_{i,j} - X_i'\hat{\beta}_j^{OLS}$ is the residuals from the regression using the data.

Planned Retirement Age Groups. Next, we include the share of individuals retiring in 6 particular age-groups, $k = \{50-54, 55, 56-59, 60, 61-64, 65\}$. Denote as $\mathcal{S}_{i,j} = (d_{i,j,1}, \ldots, d_{i,j,6})'$ the 6-element column vector of dummies where $d_{i,j,k}$ is one if member j in household i is in group k and zero otherwise. Likewise, denote $\mathcal{S}_{i,j}^{(s)}(\theta)$ as the simulated counter-part of this set of dummies. We then include as the third set of moments,

$$\mathcal{M}_3(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\begin{pmatrix} \mathcal{S}_{i,h} - \mathcal{S}_{i,h}^{(s)}(\theta) \\ \mathcal{S}_{i,w} - \mathcal{S}_{i,w}^{(s)}(\theta) \end{pmatrix}.$$

Simultaneous retirement. The final moments included relate to the retirement timing of couples. Defining the retirement calendar year as $\mathcal{C}_{i,m}$ and the simulated counterpart as $\mathcal{C}_{i,m}^{(s)}(\theta)$, the final moments are

$$\mathcal{M}_4(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{S}\sum_{s=1}^{S}\begin{pmatrix} \mathbf{1}\{\mathcal{C}_{i,h}-\mathcal{C}_{i,w}\in\{-2,-1\}\} - \mathbf{1}\{\mathcal{C}_{i,h}^{(s)}(\theta)-\mathcal{C}_{i,w}^{(s)}(\theta)\in\{-2,-1\}\} \\ \mathbf{1}\{\mathcal{C}_{i,h}-\mathcal{C}_{i,w}\in\{1,2\}\} - \mathbf{1}\{\mathcal{C}_{i,h}^{(s)}(\theta)-\mathcal{C}_{i,w}^{(s)}(\theta)\in\{1,2\}\} \\ \mathbf{1}\{\mathcal{C}_{i,h}=\mathcal{C}_{i,w}\} - \mathbf{1}\{\mathcal{C}_{i,h}^{(s)}(\theta)=\mathcal{C}_{i,w}^{(s)}(\theta)\} \end{pmatrix}.$$

Stacking all moments together gives

$$g(\theta) = (\mathcal{M}_1(\theta), \mathcal{M}_2(\theta), \mathcal{M}_3(\theta), \mathcal{M}_4(\theta))'$$

and the estimator of θ is

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

where we use as weighting a matrix, W, the inverse of the bootstrapped variances of the moments on the diagonal and zero everywhere else.

We solve the minimization problem by successively applying different minimization routines in Matlab. We perform the sequence of estimators four times and report the estimates yielding the lowest criteria function. For each of the four estimation runs, we start with MATLABs `particleswarm` which is a "global" optimization routine using randomization to search through the parameter space. We use 80 particles and switch to Nelder-Mead (`fminsearch` in MATLAB) using the best candidates from the converged particleswarm. We use $S_{sim} = 100$ simulation draws for this estimation. After the four sequences of these two algorithms, we increase the number of simulation draws to $S_{sim} = 2,000$ and do one final Nelder-Mead minimization starting at the parameters yielding the lowest objective function over the four sequences of estimators. We then report the parameter values that solves this final minimization.