Threshold effects in nonlinear models with an application to the social capital-retirement-health relationship[*]

June 2014

Abstract

This paper considers the relationship between social capital and health in the years before, at and after retirement. This adds to the current literature that only investigates this relationship in either the population as a whole or two subpopulations, pre- and post-retirement. We now investigate if there are further additional subpopulations in the years to and from retirement. We take an information criteria approach to select the optimal model of subpopulations from a full range of potential models. This approach is similar to that proposed for linear models by Gonzalo and Pitarakis (2002). Our contribution is to show how this may also be applied to nonlinear models and without the need for estimating subsequent subpopulations conditional on previous fixed subpopulations. Our main finding is that the association of social capital with health diminishes at retirement and this decreases further ten years after retirement. We find a strong positive significant association of social capital with health, although this turns negative after twenty years, indicating potential unobserved heterogeneity. The types of social capital may differ in later years (e.g. less volunteering) and hence overall social capital may have less of an influence on health in later years.
JEL Classification: I2 I1 C1

1 Introduction[1]

This paper investigates the relationship between an individual's self-assessed health (SAH) and their accumulated levels of social capital (SC), and how that relationship varies with the length of time the individual has before or after their retirement from the labour force. The rationale stems from the economic model of health production (Grossman 1972) and the developments on economics of social capital in later years (see, for example, Glaeser, Laibson, and Sacerdote 2002, Durlauf and Fafchamps 2005). The model proposed by Glaeser, Laibson, and Sacerdote (2002) states that there may be both market (e.g. work and wages) and non-market (e.g. health) returns to social capital. Investment in social capital may decline with the opportunity cost of time but increase with the occupational returns to social skills. Furthermore, investment may decline with age and their model predicts that this decline is not dependent on health, i.e. for those who are in good health, the decline still persists. Their model motivates our idea that the relationship between SAH and SC will change due to labour market status, i.e. if they retire. Glaeser, Laibson, and Sacerdote (2002) emphasise that there may be no effect of retirement on social capital investment if the exit from work is expected - in addition, there could be anticipated effects (before retirement) as well as even lagged effects (post-retirement). Following on from this, Smith (2010) proposes that exit from the labour market may impact on the potential returns to social capital, but if retirement is expected then these returns may not be discretely affected at retirement. Consequently, the effect of social capital on health may remain much the same up to the time of retirement or even at retirement. The effect in later years however may be determined by a different type of social capital.

While there has been much research on the effect of social capital on health and the effect of retirement on health, there has been less investigation into the impact of social capital on health before, at, or after retirement. There are plausible arguments as to why social capital matters for health of the older population. In terms of the effect of social capital on health, many studies have found that higher levels are associated with better self reported health for adults of all ages (see, for example, Petrou and Kupek 2008, Ronconi, Brown, and Scheffer 2010, d'Hombres, Rocco, Suhrcke, and McKee 2011). Recent studies that have specifically analysed data on older individuals include Sirven and Debrand (2008), Sirven and Debrand (2012) and van Groezen, Jadoenandansing, and Pasini (2011). Using involvement in social activities as a proxy for social capital, Sirven and Debrand (2008) find that there is a positive effect on reducing the probability of poor health and show that social participation contributes by three percentage points to the increase in the share of individuals reporting good or very good health on average. Sirven and Debrand (2012) then find that social capital does have a causal impact on health, but there was no differentiation made in this paper across subpopulations such as age groups or retirement status. van Groezen, Jadoenandansing, and Pasini (2011) find that civic participation has a positive effect on health across all European countries. In the only other study that does exist to date, to our knowledge, on the SAH-SC relationship for those in retirement, outcomes are considered only for the post-retirement periods (Smith 2010). The authors exogenously impose a discontinuity in the reponse at retirement age and find that there is a positive impact of social capital on those with poorer health at retirement. This leads us to conjecture that there could be heterogeneity in the response of SAH to SC across subpopulations before, at and after retirement, particularly if there are adverse health shocks.[2] Since we do not know precisely when anticipated or lagged effects may happen, our study is critical to further understanding of the effect of social capital on health status for older people.

In general, the retirement literature has estimated relationships pre- and post-

---

[2]Indeed this suggests that retirement could be endogenous with respect to health, if adverse health shocks causes people to retire.

retirement, which in most countries equates to pre- and post- age 65 for males and 60 for females. The SAH-SC relationship may differ depending on years to and from retirement, therefore in this paper we search for data-determined heterogeneity over the variable *years to and from retirement*. We concentrate on the model for male retirement in this paper. We follow a simple search methodology to determine the optimal number of subpopulations and their position, based on similar standard optimal model selection in linear models (Gonzalo and Pitarakis 2002). Our model allows us to estimate associations between SC and SAH.

2 Methodology

We search for sample heterogeneity in the response of SAH to variations in SC across populations defined in terms of years to retirement. We suggest a simple model selection methodology to determine both the number of subpopulations and their definition. This method is a version of latent class or finite mixture modelling, see Fruhwirth-Schnatter (2004). Our approach follows that of Gonzalo and Pitarakis (2002), who took a similar approach for linear regression models. Our approach differs from theirs in two respects. Firstly, we extend the approach to a non-linear model estimated by maximum likelihood; in our case an ordered probit model. Secondly, we estimate the definitions of all subpopulations simultaneously, rather than sequentially as they do. The details of our approach follow.

2.1 Threshold models and discrete data

Consider a parametric model for i.i.d. random vectors (Y_i, Z_i) with conditional density function $f(y|z; \theta)$ and parameter vector θ. We allow θ to take different values across a finite number of subpopulations, determined by an i.i.d. observed variable R_i which may or may not be included in Z_i. Based on observations (y_i, z_i, r_i), $i = 1, \ldots, N$, the model is assumed to take the form of the density function

$$g_M(y_i|z_i, r_i; \theta, \tau) = \sum_{m=1}^{M} f(y_i|z_i; \theta_m) \cdot r_{m,i}(\tau),\tag{1}$$

4

where $r_{m,i}(\tau) = 1\{\tau_{m-1} < r_i \leq \tau_m\}$, $1\{A\}$ is the indicator function that takes the value 1 if A is true and 0 otherwise, $\theta = (\theta_1, \ldots, \theta_M)'$, $\tau = (\tau_1, \ldots, \tau_{M-1})'$, and M is the number of subpopulations, each of which is determined by the thresholds τ_j that satisfy $-\infty = \tau_0 < \tau_1 < \ldots < \tau_{M-1} < \tau_M = \infty$. These thresholds are not assumed to be known and will be estimated from the data, along with M. Thus $r_{m,i}$ are dummy variables, each taking the value 1 if observation i is classified into subpopulation m based on thresholds τ, and 0 otherwise. For example, if $M = 2$ then the model is

$$g_2(y_i|z_i, r_i; \theta_1, \theta_2, \tau_1) = \begin{cases} f(y_i|z_i; \theta_1), & \text{if } r_i \leq \tau_1, \\ f(y_i|z_i; \theta_2), & \text{if } r_i > \tau_1. \end{cases}$$

An implicit assumption in specification (1) is that the distribution of Y_i, once conditioned on Z_i, depends on R_i only through the determination of the subpopulation. If R_i is included in Z_i then this assumption is automatically satisfied. If R_i is not included in Z_i then this assumption need not hold, although as a practical matter it may be appropriate in this case to expand the model in (1) to $f(y_i|z_i, r_i; \theta_j, \eta_j)$ in order to exploit any additional explanatory power in R_i.

If the number of subpopulations M is known then estimation of the remaining parameters in (1) can be achieved by maximising the log-likelihood

$$\ell_M(\theta, \tau) = \sum_{i=1}^{N} \ln g_M(y_i|z_i, r_i; \theta, \tau) \tag{2}$$

over θ and τ, yielding estimators $\hat{\theta}^{(M)}$ and $\hat{\tau}^{(M)}$ and maximised log-likelihood value $\ell_M\left(\hat{\theta}^{(M)}, \hat{\tau}^{(M)}\right) = \hat{\ell}_M$. The estimation of θ follows by applying maximum likelihood separately to each of the M sub-samples specified by the thresholds τ. These M maximum likelihood estimators are denoted $\hat{\theta}^{(M)}(\tau) = \left(\hat{\theta}_1^{(M)}(\tau), \ldots, \hat{\theta}_M^{(M)}(\tau)\right)'$. The concentrated log-likelihood is then $\ell_M\left(\hat{\theta}^{(M)}(\tau), \tau\right)$, where conditional independence of observations is assumed. The practical estimation of the threshold model involves searching over τ values by grid search within the support of $(r_1, \ldots, r_N)$, with M maximum likelihood estimators computed for each τ tried. This procedure is generally practical for small numbers of subpopulations, say 2, 3, 4, 5, but then quickly becomes computationally prohibitive as the dimension of M grows.

Following the suggestion of Gonzalo and Pitarakis (2002) in linear regression models, we select M on the basis of an information criterion; that is

$$\hat{M} = \arg\min_{M} IC(M),$$

where:

$$IC(M) = -2\hat{\ell}_M + \lambda_N p_M; \text{ where } M = 1, ..., M^* \tag{3}$$

M^* is the maximum number of subpopulations to be considered for the analysis. This may be specified on a theoretical basis, but more commonly will be specified on the practical grounds of the maximum number of subpopulations for which estimation is computationally feasible. λ_N is a deterministic function of N; and p_M is the number of parameters in a model with M subpopulations. If θ_j contains k parameters then the model contains Mk parameters in θ and $M - 1$ parameters in τ, so that

$$p_M = Mk + M - 1.$$

Some common choices of λ_N include the following.

$$
\begin{array}{ll}
\lambda_N = \ln N & BIC/SC \text{ (Schwarz 1978)} \\
\lambda_N = 2 & AIC \text{ (Akaike 1987)} \\
\lambda_N = 1 + \ln N & CAIC \text{ (Bozdogan 1987)} \\
\lambda_N = 2\ln\ln N & HQIC \text{ (Hannan and Quinn 1979)}
\end{array}
$$

These criteria are derived from differing principles and have differing properties. The BIC, $CAIC$ and $HQIC$ penalties can be shown to result in consistent model selection for the true M, while the AIC has some non-zero probability of over-selection of M asymptotically (Gonzalo and Pitarakis 2002).[3]

In summary what we propose is to estimate via a grid search, all possible $m = 1, \ldots, M$ models, and choose the one which minimises the appropriate IC.[4]

---

[3]Simple analysis of λ_N shows that, for $N > 8$, $\hat{M}_{AIC} \geq \hat{M}_{HQ} \geq \hat{M}_{BIC} \geq \hat{M}_{CAIC}$.

[4]*Stata* code for the search procedures employed in this paper is available from the authors on request.

2.2 The Ordered Probit model

We consider an ordered probit (OP) model,[5] derived from a latent variable Y_i^* of the form

$$Y_i^* = Z_i'\gamma + U_i, \tag{4}$$

with Z_i being a vector of explanatory variables (with no constant term), γ a vector of unknown coefficients and U_i a standard normal disturbance term. This latent variable translates to the observed ordered, discrete, random variable Y_i, with $j = 1, ..., J$ outcomes, via the mapping

$$Y_i = j \cdot 1\left\{\mu_{j-1} < Y_i^* \leq \mu_j\right\},$$

where $-\infty = \mu_0 < \mu_1 < \ldots < \mu_{J-1} < \mu_J = \infty$ are the boundary parameters with $\mu = \left(\mu_1, \ldots, \mu_{J-1}\right)'$ to be estimated in addition to γ. With $\theta = (\gamma, \mu)$, the OP density function is given by

$$f\left(y_i | z_i; \theta\right) = \Phi\left(\mu_{y_i} - z_i'\gamma\right) - \Phi\left(\mu_{y_i-1} - z_i'\gamma\right).$$

Following the general model (1) we define, for each subpopulation m, coefficient vectors γ_m and boundary parameters $\mu_m = \left(\mu_{m,1}, \ldots, \mu_{m,J-1}\right)'$ for $m = 1, \ldots, M$, so that an M-subpopulation OP model with threshold variable r_i has conditional density function

$$g_M\left(y_i | z_i, r_i; \theta, \tau\right) = \sum_{m=1}^{M} \left(\Phi\left(\mu_{m,y_i} - z_i'\gamma_m\right) - \Phi\left(\mu_{m,y_i-1} - z_i'\gamma_m\right)\right) \cdot r_{m,i}\left(\tau\right),$$

where $\theta = \left(\gamma_1, \ldots, \gamma_M, \mu_1, \ldots, \mu_{M-1}\right)$. The log-likelihood is then computed using equation (2) and, if each γ_m contains k coefficients, the penalty term p_M in (3) is

$$p_M = Mk + M\left(J - 1\right) + \left(M - 1\right).$$

The computation of $\hat{M}$ can proceed using any IC in (3).

A model in which γ differs across subpopulations, but the boundary parameters μ do not, can be specified as

$$g_M\left(y_i | z_i, r_i; \theta, \tau\right) = \sum_{m=1}^{M} \left(\Phi\left(\mu_{y_i} - z_i'\gamma_m\right) - \Phi\left(\mu_{y_i-1} - z_i'\gamma_m\right)\right) \cdot r_{m,i}\left(\tau\right),$$

---

[5]The proposed methodology can similarly be applied to any nonlinear model estimated by maximum likelihood, where the relevant density is entered into equation (2).

where now $\theta = (\gamma_1, \ldots, \gamma_M, \mu)$, and the IC penalty p_M is

$$p_M = Mk + (J - 1) + (M - 1).$$

We assume, as in our application below, that only a subset of the γ coefficients vary across subpopulations. Let $z_{1,i}$ be the k_1 regressors whose coefficients, denoted $\gamma_{1,1}, \ldots, \gamma_{1,M}$, vary across subpopulations and let the remaining k_2 regressors $z_{2,i}$ have constant coefficients γ_2. In this case the conditional density simplifies to

$$g_M \left(y_i | z_i, r_i; \theta, \tau \right) = \sum_{m=1}^{M} \left(\Phi \left(\mu_{y_i} - \left(z_{1,i}' \gamma_{1,m} + z_{2,i}' \gamma_2 \right) \right) - \Phi \left(\mu_{y_i - 1} - \left(z_{1,i}' \gamma_{1,m} + z_{2,i}' \gamma_2 \right) \right) \right) \cdot r_{m,i} \left(\tau \right),$$

where $\theta = \left(\gamma_{1,1}, \ldots, \gamma_{1,M}, \gamma_2, \mu \right)$, with IC penalty

$$p_M = Mk_1 + k_2 + (J - 1) + (M - 1).$$

Indeed, our methodology applies even if one has no prior information regarding whether all, or subsets of, the parameters of the model are constant or vary by subpopulation. All competing models can be simply compared using their IC.

Model selection based on IC can be given a hypothesis testing interpretation (Gonzalo and Pitarakis 2002). If a model with M_1 subpopulations is preferred by an $IC(M)$ to a model with M_2 subpopulations, where $M_1 > M_2$, the ordering $IC(M_1) < IC(M_2)$ implies that $2 \left(\hat{\ell}_{M_1} - \hat{\ell}_{M_2} \right) > \lambda_N (p_{M_1} - p_{M_2})$. That is, the smaller model M_2 would be rejected in favour of the larger model M_1 by a likelihood ratio (LR) test using a non-standard critical value of the form $\lambda_N (p_{M_1} - p_{M_2})$. The level of significance of this likelihood ratio test depends on the choice of λ_N, with larger λ_N (*e.g.*. BIC and $CAIC$) implying a test with lower size and power, hence preferring smaller models. The preceding argument also holds for $M_1 < M_2$ since then $2 \left(\hat{\ell}_{M_2} - \hat{\ell}_{M_1} \right) < \lambda_N (p_{M_2} - p_{M_1})$, implying the smaller model M_1 is not rejected in favour of the larger model M_2.

We note here that our method is explicitly based upon certain identifying assumptions and restrictions which may be an issue for researchers applying similar techniques elsewhere. Explicitly these are:

1. The data we have to hand is cross-sectional, so that we cannot easily condition on any individual unobserved heterogeneity as would be possible in a panel

data setting. We do note, however, that our suggested techniques are similarly appropriate to the use of nonlinear panel data models, one simply has to adjust the likelihood functions accordingly (see, for example, Greene 2010, Wooldridge 2010).

2. The model does not contain any endogenous regressors.

3. Any potential endogeneity of the threshold variable - *years to and from retirement* - is not dealt with in our paper.

4. There are no omitted variables related to the variation on the response parameter.

5. Variation in responses across subpopulations is deterministic and variation within each range is homogeneous.

3 Data

We use data from Wave 1 of the SHARE (Survey of Health, Ageing, and Retirement) data set, collected in 2004/2005. This is a cross-national individual level longitudinal data set including more than 38,000 Europeans aged over 50. The database contains rich information on health status, socioeconomic characteristics, housing, as well as various proxies for social capital. Data are available across 11 countries (Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Greece, Switzerland, Israel) in which we have self-reported measures of health, various indicators of social capital e.g. volunteering, (that we later utilise to define a composite measure of social capital), as well as a range of personal demographics typically used in this literature (such as education levels and a full range of country dummies). These data have been widely used before and in related contexts (see, for example, Sirven and Debrand 2012)**.**

The variable we wish to model is an individual's self-assessed health, denoted SAH_i, which is defined on a five point Likert-scale:

$$SAH_i = \begin{cases} 1 & \text{poor} \\ 2 & \text{fair} \\ 3 & \text{good} \\ 4 & \text{very good} \\ 5 & \text{excellent} \end{cases}.$$

The social capital variable, denoted SC_i, is measured as a continuous index, increasing in levels of social capital and incorporating many facets of social capital (including volunteering, caring, networking, family ties). This index has been developed in detail and discussed in depth by Gannon and Roberts (2012). However, in short, the index was found to be a good and robust measurement of social capital and predictor of health using several definitions of self assessed health and mental health. The index is scaled from 0 to 1 and has a mean score of 0.38. The SAH variable has a mean value of 3.24.

The relationship between SAH_i and SC_i is hypothesised to vary with the years to and from retirement variable, R_i. This variable is constructed by us, for all retirees, from a question within SHARE, "In which year did your last job end?". In Wave 1 of the data, most retired respondents answered this question, whereas in Wave 2 there were many missing values.[6] We therefore concentrate on Wave 1 of the data only. For employees, we calculated years to retirement based on expected retirement age for each country. In most countries this was age 65 for men and age 60 for women; although here we only consider our model for males. The R_i variable therefore takes on 40 possible values, ranging from -15 years to retirement to 25 years past retirement (we only analyse individuals to age 80 due to the very small sample sizes beyond this, hence our maximum number of years past retirement is 25). For example, if a person is aged 55 and we expect them to retire at age 65, this variable will take the value of -10. On the other hand, if they are aged 75 and retired at age 65, then this variable takes a value of 10.

Following the literature, the model also includes an additional set of control variables, standard demographics, such as log of household income, highest education level, married/living with partner, and country dummies, that are included in a regressor vector (net of SC_i) X_i. Our estimation sample size is $n = 10,320$.

---

[6]We have checked the response rate for those who were present in wave 1 and find that even with matching up responses across waves, there is still 12% missing on this variable.

4 Empirical model

The standard (constant-coefficient) OP model for the SAH-SC relationship is based on the latent variable representation

$$SAH_i^* = \beta SC_i + \alpha' X_i + U_i, \tag{5}$$

where, with regard to the earlier notation, $Z_i = (SC_i, X_i)$, $U_i \sim$ i.i.d.$N(0,1)$ and SAH_i^* is an unobserved continuous measure of *latent* health. As is usual, and as described above, this latent variable is assumed to be mapped into the observed dependent variable SAH_i by

$$SAH_i = j \cdot 1\left\{\mu_{j-1} < SAH_i^* \leq \mu_j\right\}, \quad j = 1, \dots, 5, \tag{6}$$

where $\mu_1, \dots, \mu_4$ are the boundary parameters to be estimated (setting $\mu_0 = -\infty$ and $\mu_5 = \infty$).

We now extend this constant-coefficient model to allow for the hypothesised differential effect of social capital on health with respect to an individual's position with regard to time to and past retirement. Thus our threshold model now allows the coefficient β on SC_i to vary according to years to and from retirement, while assuming that α and $\mu_1, \dots, \mu_4$ are constant across all individuals.[7] This threshold model takes the general form

$$SAH_i^* = \sum_{m=1}^{M} \beta_m \left(SC_i \cdot R_{m,i}\right) + \alpha' X_i + U_i, \tag{7}$$

where the mapping of equation (6) is unchanged and the subpopulation indicators are defined as

$$R_{m,i} = 1\left\{\tau_{m-1} < R_i \leq \tau_m\right\}, \quad m = 1, \dots, M,$$

with the number of subpopulations M and the threshold parameters $\tau_1, \dots, \tau_{M-1}$ (with $\tau_0 = -\infty$, $\tau_M = \infty$) to be estimated from the data. This model splits the data into M subpopulations, defined by $R_{1,i}, \dots, R_{M,i}$, that each have a constant

---

[7]A more general model would allow all parameters to so vary. However, we have no strong priors as to why this would be the case. Moreover, as described above, such a case could still easily be entertained by the methods suggested.

coefficient on SC_i given by $\beta_1, \ldots, \beta_M$. Setting $M = 1$ in (7) gives the constant coefficient model (5) as a special case (in which case $\tau_0 = -\infty$, $\tau_1 = \infty$ and $R_{1,i} = 1$ for all i).

In this particular application, where the threshold variable R_i takes one of only 40 values (due to the restrictions noted above) and the sample size is relatively large, it is possible to set $M = 40$ in equation (7) and re-define the threshold indicators as

$$R_{m,i} = 1\left\{R_i = m\right\}, \quad m = 1, \ldots, 40,$$

in the process removing the need for explicit estimation of the threshold parameters. That is, as we have a discrete threshold variable, and multiple observations per observed category, it is possible to estimate the fully unrestricted model with $M = 40$ subpopulations.[8] Clearly this yields a very highly parameterised model (with a different coefficient on SC_i for each different value of R_i), which although allowing for a very flexible SAH-SC profile across all years to and from retirement, is unlikely to be entertained as an appropriate empirical model. Such an unrestricted model is only possible when the threshold variable takes a small number of possible values (relative to the sample size). However, in this instance it provides a useful model for comparative purposes.

In summary, we can consider three benchmark models involving no search procedures. Firstly, where the β coefficient is fixed. Secondly, where it is allowed to vary for all observed values of R_i. And finally, where we adopt the approach in the literature to essentially translate the problem into a binary one, such we simply split SC_i into two: pre- and post-retirement.[9] Compared to these benchmark models, we then consider a range of models which entertain all possible combinations of $m = 1, \ldots, M^* = 5$ subpopulation models (where M^* is the maximum set value of M), and optimally choose M according to the IC stated above. Note that, in practical terms, the researcher must pick a value for M^*. Ideally this should be as

---

[8]This would not be possible if the threshold variable were continuous, and moreover is not one that many applied researchers would necessarily entertain, being particularly unparsimonious.

[9]An anonymous referee suggested we should include age squared and perhaps an interaction between SC and age, and present this as our benchmark model. They also suggested presenting results as partial effects. However, we estimated these models and they did not add any useful information to the benchmark model. In addition, the partial effects demonstrate a very similar pattern to those of the coefficients, so we excluded these for reasons of space.

large a number as possible (noting the trade-off with regard to additional estimation time). We also note that in practical terms, the procedure can be sped-up by not searching over the full support of R_i. For example, in our application, it is rather unlikely that any threshold will occur at either end of the distribution of R_i.[10]

5 Results

Although not reported for reasons of space, a limited set of Monte Carlo experiments was considered to evaluate the performance of our techniques.[11] All of the techniques showed good performance in correctly selecting the number and position of the thresholds; and fared better the larger the difference across subpopulations and with fewer subpopulations. However, these did reveal that the BIC would be the preferred criterion across a range of scenarios considered. Thus, in general, we would recommend the use of BIC in such a search procedure. This is in line with the findings of (Gonzalo and Pitarakis 2002) (p. 337) for the linear model.

Turning now to the empirical application, the results for the benchmark models (*Imposed Subpopulations*), as well as the optimal model for each $M = 1, \ldots, 5$ (*Estimated Subpopulations*) are presented in the form of estimated coefficients in Table 1.[12] Table 1 also presents the various maximised likelihood functions and information criteria along with the optimally chosen candidates for τ in each of these.

Firstly, consider the exogenously imposed number of subpopulation results (columns

---

[10]We note here the possibility of the *years to and from retirement* variable being potentially endogenous, and also that for those yet to retire, that this variable has been measured with some noise. However, rigorous treatment of either these aspects is beyond the scope of the current paper, and moreover the consequences of such are not yet known. For the majority of those in our sample who are retired, we know the retirement year for all, except for 3% whose information is missing from the sample. We treat this as an accurate date of retirement. For those not retired, we estimate the years to retirement based on the statutory date of retirement and acknowledge that this is measured with noise.

[11]The full set of these results are available from the authors on request.

[12]For reasons of space we only present the coefficients for the social capital variables. Full coefficient results are available on request, but are in general accordance with those expected and found in the previous literature.

$5-7$). The constant coefficient model ($M = 1$) in the fifth column of results, would correspond to the situation where the researcher did not impose any threshold in the SAH-SC relationship with regard to years to and from retirement. The results show the expected significant positive relationship between SAH and SC, with $\hat{\beta}_1 = 0.457$. In light of previous literature, we then allow for an exogenously determined threshold in this relationship at the point of retirement and obtain the results in column 6. Here we indeed see that the assumption of a constant relationship looks questionable, with a halving of the coefficient from 0.638 to 0.305 post-retirement. Again, as expected the effect of SC on health diminishes after retirement. Although unlikely to be considered as a sensible, or parsimonious, model by most researchers, the final column gives the results of allowing a different effect of SC on SAH for every observed value of time to and from retirement. Note that for reasons of space, we do not report all these coefficients, but they are presented and discussed below in Figure 1. A comparison of these exogenously imposed subpopulations, across all IC measures, clearly shows that the $M = 40$ model is heavily over-parameterised, whereas exogenously imposing a single threshold at retirement does appear warranted.

We now turn to the optimally chosen threshold models. Although our Monte Carlo simulations suggested a preference for the BIC (especially in comparison to the AIC), here this is inconsequential, as all methods choose the same number of subpopulations and at the same places for all possible candidates for all τ. Thus from Table 1, we see that for all possible $M = 2$ candidate models, we find that the optimal one corresponds to the single threshold, eight years after the year of retirement. Allowing for one more subpopulation ($M = 3$) now finds a threshold at retirement and then another one several (seventeen) years past retirement. Moving to $M = 4$, we find thresholds at retirement, an intermediate threshold ten years past retirement, and a considerably later threshold at 19 years past retirement. Finally, we consider all possible $M = 5$ candidate models, although here this appeared to be over-parameterised as the $M = 4$ model was found to be preferable, as we now discuss.

Before analysing the results in more detail, it is useful to compare IC measures for all considered models. It is clear that all of the four IC measures favour the

optimally chosen $M = 4$ model, with thresholds at retirement, and 10 and 19 years past retirement. Thus the results for $M = 4$ show that we do, indeed, find a threshold in the social capital/health relationship in the year at retirement ($\tau_1 = 0$). This is an interesting finding and confirms the existing results in the literature which simply exogenously impose a threshold here. Our main finding is that the effect of SC on health diminishes at retirement and additionally the effect decreases further in later years after retirement. This could reflect lagged effects and the fact that the types of social capital may differ in later years (e.g. less volunteering) and hence overall social capital may have less of an influence on health. Nonetheless, there is still a strong significant effect of social capital on health for all individuals up to twenty years post-retirement, potentially reflecting the impact of both current and lagged effects at older age.

Our results also show that the pre-retirement relationship does not seem to be subject to any threshold, and hence across all ages from 50 upwards, the effect of social capital on health is relevant but, in contrast to the post-retirement era, it is not overly important to distinguish the years to retirement. While research to date has only specified retirement status as a binary outcome, we now show the importance of specifying years since retirement.

Prior to the first threshold, social capital is found to have a significant positive effect on health $\left(\hat{\beta}_1 = 0.696\right)$. In the second subpopulation, from the year after retirement until ten years after retirement, this SC effect remains positive but is weaker than before retirement $\left(\hat{\beta}_2 = 0.366\right)$. The third subpopulation, from 11-19 years after retirement, essentially shows that SC here has only a very weak effect on health levels ($\hat{\beta}_3 = 0.063$, and insignificant); whereas the final fourth subpopulation, from 20 years after retirement onwards, shows a significantly negative SAH-SC relationship $\left(\hat{\beta}_4 = -0.551\right)$.

These results compare to previous research, whereby Smith (2010) looked at post-retirement health status conditioned on health status at retirement year, and found a positive link between SC and health for those with baseline poor health. That paper however, did not look at years since retirement. Our results suggest that exogenously imposing a single threshold at year of retirement, could yield potentially

biased and misleading results and policy inference. While we find that social capital does impact on health at retirement, we also find large significant effects in later years, even if at a smaller magnitude. Similar to Rocco and Suhrcke (2013) we find social capital is good for health pre- and post-retirement (unless one is in retirement at least 19 years) and while we also can only propose limited policy implications, these should nonetheless be considered. For example, Rocco and Suhrcke (2013) propose interventions such as provision of support or promotion of volunteering. In our case, this may be important in years after retirement, as well as in the years leading to retirement or at retirement.

The estimated weakening relationship between social capital and health in later years suggests it is likely that the *type* of social capital is relevant here, and that high levels of SC in earlier years are likely to reflect higher levels of volunteerism, in later years family ties and networks may be more dominant but these do not have as great an association with health production (Gannon and Roberts 2012).

In Figure 1 we present a range of coefficient estimates for the SC variable by years to and from retirement. We consider a range of models here: *General*, where, as described above, we have a separate effect for each observed realisation of R_i; *Base*, where there is no threshold effect; *Optimal* - which is the optimal threshold model (using any IC); and *Pre and Post*, which would be the effect of simply splitting the SC index at retirement age, as would typically be done in the literature.

Considering the parameters first, and starting from the most rigid approach with regard to years to and from retirement, the static positive effect of SC on health is exhibited by the horizontal *Base* line. Allowing for a threshold, exogenously determined, at retirement age, clearly suggests that the effect of SC, while remaining positive, diminishes for retirees. However, now consider the results for the fully unrestricted model, with $M = 40$ and coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_{40}$. These clearly suggest the SAH-SC relationship may be fairly constant before retirement, but then slopes downwards, becoming less positive, and even negative a number of years after retirement. Compare this to the optimally determined relationship, which again shows a constant effect prior to retirement, but then one that diminishes in effect, and indeed becomes negative at 20 years past retirement. One good interpretation of our

16

Table 1: Threshold Model Estimation Results; Standard Errors in Parentheses

	Estimated Subpopulations				Imposed Subpopulations		
Column	1	2	**3**	4	5	6	7
Number of subpopulations	2	3	4	5	1	2	40
Log-likelihood $\hat{\ell}_M$	-13954	-13942	$\mathbf{-13936}$	-13934	-13969	-13956	-13922
p_M	21	22	**23**	24	20	21	59
$BIC\,(M)$	28103	28334	**28084**	28090	28123	28107	28398
$AIC\,(M)$	27951	28174	**27918**	27916	27979	27955	27963
$CAIC\,(M)$	28124	28356	**28107**	28114	28143	28128	28458
$HQIC\,(M)$	28002	28228	**27974**	27975	28027	28006	28110
Threshold parameters							
τ_1	8	0	**0**	-8	n/a	0	-
τ_2		17	**10**	0			-
τ_3			**19**	10			-
τ_4				19			
Social Capital coefficients							
$\hat{\beta}_1$	0.551** (0.072)	0.670** (0.078)	0.696** (0.078)	0.621** (0.090)	0.457** (0.694)	0.638** (0.078)	-
$\hat{\beta}_2$	0.165* (0.087)	0.318** (0.075)	0.366** (0.076)	0.764** (0.087)		0.305** (0.076)	-
$\hat{\beta}_3$		-0.294^* (0.136)	0.063 (0.136)	0.374** (0.077)			-
$\hat{\beta}_4$			-0.551^{**} (0.007)	0.094 (0.099)			-
$\hat{\beta}_5$				-0.509^{**} (0.171)			

Notes: standard set of demographics (education level, age in years, marital status, log of household income) and country dummies included. ** and * respectively denote significance at 5 and 10%. For all M all IC measures selected the same model. τ values denote years past retirement. τ and β not presented for $M = 40$ model for reasons of space. The estimated subpopulation results for each M correspond to the optimal model for all $m = 1, \ldots, M$ models considered. Overall preferred results in **bold**.
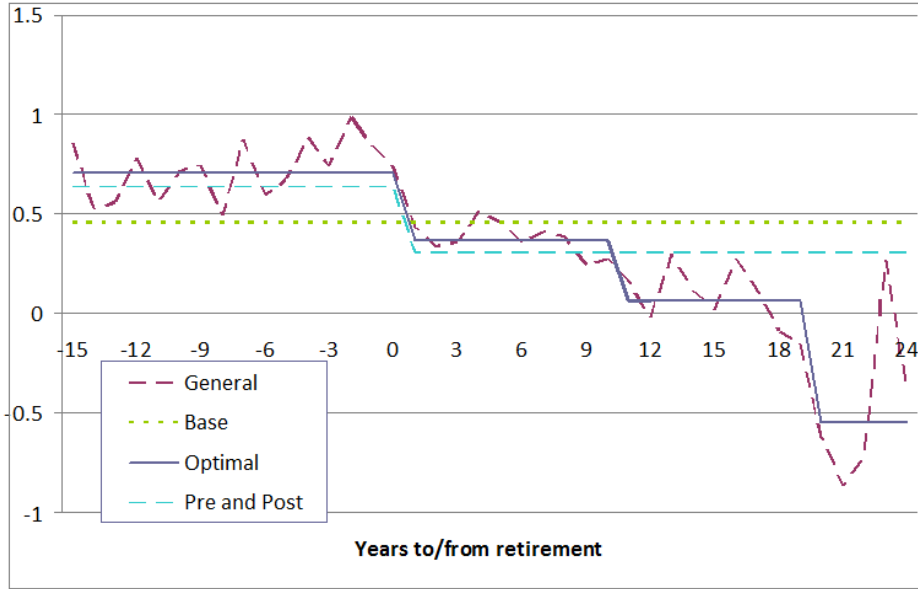
Figure 1: Estimated coefficients under alternative model restrictions

threshold model is that it captures the most flexible relationship in a parsimonious way by estimating a much smaller number of subpopulations, within each of which the relationship is constant. Indeed, it is exactly in this manner that it is possible to identify both the position, and form, of any thresholds in the SAH-SC relationship with regard to years to and from retirement. Although not reported for reasons of space, as would be expected the results for the partial effects simply reflect those patterns in the coefficients values.

We are reticent to label any of these effects as *causal*, as there is the distinct possibility of reverse causation here; we therefore label the estimated relationships as associative. Furthermore, contemporaneous shocks may affect health and social interaction and a change in health could be affected by unobservables that correlate with social connectedness at baseline (Smith 2010).

6 Concluding comments

In this paper we considered heterogeneity in the response of health to social capital, pre- and post-retirement. It is likely that the composition and level of social capital will change past retirement, such that it is likely to have differing effects on health

in this period. The results showed that by estimating these important relationships over a range of years, as opposed to simply pre- and post-events, we revealed new results of policy relevance. Indeed, the results suggested that exogenously imposing a single threshold at retirement age could both give misleading policy advice and disguise potentially interesting results.

Our suggested technique was to estimate all possible subpopulations, and to pick that which minimises the *BIC*, as evidence suggested this out-performed other available *IC* metrics. We suggest that the technique is generally applicable to any discrete choice, or limited dependent variable, model estimated by maximum likelihood techniques, and thus would be applicable in a vast array of instances. Indeed, we envisage a wide array of potential applications in the health economics field, such as nonlinear effects of income on health behaviours; of body mass index on health outcomes; of child's age and parental income in models of child health; of years to and from retirement in many models of health behaviours.

References

AKAIKE, H. (1987): "Information Measures and Model Selection," *International Statistical Institute*, 44, 277–291.

BOZDOGAN, H. (1987): "Model Selection and Akaike's Information Criteria (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52, 345–370.

D'HOMBRES, B., L. ROCCO, M. SUHRCKE, AND M. MCKEE (2011): "Does social capital determine health:evidence from eight transition countries," *Health Economics*, 19(1), 56–74.

DURLAUF, S., AND M. FAFCHAMPS (2005): "Social Capital," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1, chap. 26, pp. 1639–1699. Elsevier, 1 edn.

GANNON, B., AND J. ROBERTS (2012): "Social Capital: Bridging the Theory and Empircal Divide," Discussion paper, Working Paper: Universities of Leeds and Sheffield.

GLAESER, E., D. LAIBSON, AND B. SACERDOTE (2002): "An Economic Approach to Social Capital," *Economic Journal*, 112(483), F437–F458.

GONZALO, J., AND J.-Y. PITARAKIS (2002): "Estimation and Model Selection Based Inference in Single and Multiple Threshold Models," *Journal of Econometrics*, 110, 319–352.

GREENE, W. (2010): "Testing hypotheses about interaction terms in nonlinear models," *Economics Letters*, 107(2), 291–296.

GROSSMAN, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223–255.

HANNAN, E., AND B. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, B*, 41, 190–195.

PETROU, S., AND E. KUPEK (2008): "Social capital and its relationship with measures of health status: evidence from the Health Survey for England," *Health Economics Letters*, 17(1), 127–143.

ROCCO, L. FUMAGALLI, E., AND M. SUHRCKE (2013): "From social capital to health - and back," *Health Economics*, DOI:10.1002/hec.2934.

RONCONI, L., T. BROWN, AND R. SCHEFFER (2010): "Social Capital and Self-Rated Health in Argentina," *Health Economics*, 21(2), 201–208.

SCHWARZ, G. (1978): "Estimating the Dimensions of a Model," *Annals of Statistics*, 6(2), 461–464.

SIRVEN, N., AND T. DEBRAND (2008): "Social participation and healthy ageing: an international comparison using SHARE data.," *Social Science and Medicine*, 65, 842–854.

———— (2012): "Social capital and health of older Europeans: Causal pathways and health inequalities," *Social Science and Medicine*, 75(7), 1288–1295.

SMITH, S. (2010): "Social Connectedness and Retirement," Discussion paper, CMPO Working Paper No. 10/255.

VAN GROEZEN, B., R. JADOENANDANSING, AND G. PASINI (2011): "Social capital and health across European countries," *Applied Economics Letters*, 18(12), 1167–1170.

WOOLDRIDGE, J. (2010): *Econometric Analysis of Cross Section and Panel Data 2e*. MIT Press.