ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Extending Unobserved Heterogeneity – A Strategy for Accounting for Respondent Perceptions in the Absence of Suitable Data

Timothy A. Weterings, Mark N. Harris and Bruce Hollingsworth

May 2012

Working Paper 12/12

Extending Unobserved Heterogeneity - A Strategy for Accounting for Respondent Perceptions in the Absence of Suitable Data

Timothy A. Weterings[1]
Department of Econometrics and Business Statistics, Monash University, Clayton, Melbourne

Mark N. Harris
Curtin Business School, Curtin University, Perth

Bruce Hollingsworth
Division of Health Research, Lancaster University, UK

Summary: This research proposes that, in cases where threshold covariates are either unavailable or difficult to observe, practitioners should treat these characteristics as latent, and use simulated maximum likelihood techniques to control for them. Two econometric frameworks for doing so in a more flexible manner are proposed. The finite sample performance of these new specifications are investigated with the use of Monte Carlo simulation. Applications of successively more flexible models are then given, with extensive post-estimation analysis utilised to better assess the likely implications of model choice on conclusions made in empirical research.

Keywords: Ordered Choice Modeling, Unobserved Heterogeneity, Simulated Maximum Likelihood
JEL classification: C01, C23, C52

[1]Corresponding author contact details:
Email: Timothy.weterings@monash.edu
Ph: +61 9902 0886

1 Introduction

A large amount of literature recognises the need for extensions and for methodologies that relax the restrictive assumptions required for the standard Ordered Probit model. Since (Terza, 1985), many authors have attempted to account for additional heterogeneity through the use of threshold covariates (see, for example, Pudney and Shields (2000), Boes and Winklemann (2006), Pfeifer and Cornelissen (2010), Garcia-Gomez, Jones, and Rice (2010) and many others). These applications attempt to make the thresholds more flexible, and account for differences in how individuals categorise their latent propensities. The result is often better overall model fit; yet without the use of exclusive variables in the thresholds and latent regression, true directional effects of the latent regression variables on the underlying utility function cannot be determined (Weterings, Greene, Harris, and Hollingsworth, 2011). That is, due to a lack of separability of one (usually the first) threshold and the latent regression, these variables must instead be interpreted as relative to a particular threshold, so the true directional effects of threshold covariates assumed to exist in the underlying data generating processes are lost. Refer to Weterings et al. (2011) for further details.

One way to to deal with this lack of separability is to make assumptions about the relationships between the covariates in different parts of the ordered choice model. Specifically, it would be useful for the analyst to have variables in the latent regression that are expected to only affect the latent variable, and variables in the thresholds that are expected to only affect the thresholds. However, in many applications, such variables are hard to find.

An alternative to the search for threshold variables that is growing in popularity is the use of vignette questions to help anchor the variation in the thresholds that is attributable to differences in respondent perceptions (King, Murray, Salomon, and Tandon, 2004). However, vignette questions present the issue of a lack of availability. Collecting answers to an appropriate number of vignette questions is time consuming and may lead to set avoidance later in the survey if used excessively. In addition, the fact that vignette questions must be designed to help identify respondent views on a particular issue means that mainstream use in large surveys is likely to be limited at best, due to the increased cost of collecting the data.

Another approach to accounting for threshold heterogeneity, that draws on work done by Cameron and Heckman (1998), and discussed in Greene and Hensher (2010), is to treat the thresholds as though they contain an unobserved (stochastic) component. Such an approach requires either theoretical or parametric restrictions in order to guarantee the ordering of the thresholds, with examples of such given in Cunha and Navarro (2006) and also postulated in Greene and Hensher (2010), the latter outlining a simulated maximum likelihood approach that is reasonably easy to implement. Due to the lack of required threshold covariates, this approach might be seen as a desirable route to effectively separating the effect of threshold heterogeneity from the latent regression.

An issue with the existing unobserved threshold heterogeneity methodologies, however, is the use of possibly restrictive assumptions about relationships between the threshold covariates when these covariates are treated as unobservable. In order to address this, we propose a simulated maximum likelihood methodology to draw correlations between the unobserved components of the thresholds. This is intended to aid model flexibility. This extended model is further built upon by allowing a heteroscedastic variance, which we treat as unobservable in this application, and followed by the like extensions of correlations between the stochastic thresholds and stochastic variances. As far as we are aware, such extensions have not been discussed so far within the literature. Each of these models is considered within a panel data context, ensuring relevance to most practitioners using large datasets, and helping to identify these unobserved effects.

2 Econometric Framework

The modeling framework utilised for discrete, ordered outcomes extends from the Ordered Probit model. This framework is introduced in panel data form, to increase relevance to practitioners, for whom the issue of unsuitable threshold covariates is likely to be pertinent. As is usual, we assume there to be some latent propensity, y^*_{it}, which is a linear function of observable covariates, x_{it}, a time-constant individual error component, α_i, and a (normally distributed) idiosyncratic error term, ϵ_{it}.

$$y^*_{it} = x'_{it}\beta + \sigma_\alpha \alpha_i + \epsilon_{it}$$

where it is assumed that $\alpha_i \sim i.i.d\ N(0,1)$. Thresholds, μ_{ij}, denote the level of underlying utility required for each outcome to be selected, and are assumed to vary across individuals but not across time for identification purposes. Assuming J possible outcomes, the implied probabilities are:

$$Pr(y_{it} = j|x_{it}) = \Phi(\mu_{ij} - x'_{it}\beta) - \Phi(\mu_{i,j-1} - x'_{it}\beta), \quad j = 1, \ldots, J$$

where J is the number of distinct outcomes of the observable variable, y_{it}. The restrictions of $\mu_{i0} = -\infty$, $\mu_{i1} = 0$, $\mu_{iJ} = \infty$, and $Var(\epsilon_{it}) = 1$ apply for identification.

2.1 The Individual Threshold Specification

In the case that the analyst is able to observe some of the factors that potentially influence the thresholds, the thresholds may be functions of both deterministic and stochastic components. The primary scenario that this research is concerned with is that where observed threshold-identifying covariates are unavailable – a common case for much empirical research. As a result, threshold variation is assumed to be accounted for completely by the unobservable components, ω_{ij}. To ensure the ordering of the thresholds, a Hierarchical Ordered Probit (HOPIT) specification is applied, as per Greene and Hensher (2010). This helps to ease issues of identification and zero/negative probabilities that may arise under less strict functional forms.

$$\mu_{i1} = \lambda_1 + \eta_1 \omega_{i1}$$

$$\mu_{ij} = \mu_{i,j-1} + e^{(\lambda_j + \eta_j \omega_{ij})}, \;\; 2 \leq j \leq J-1$$

$$\mu_{i0} = -\infty, \mu_{iJ} = \infty$$

Outcomes for the observable variable, y_{it}, are thus observed on the criteria $y_{it} = j$ if $\mu_{i,j-1} \leq y^*_{it} < \mu_{ij}$. For this study, issues of separability of individual effects in the latent regression and the (linear) first threshold are resolved by collecting the latent regression individual effect in the equation of the first threshold. This results in the first threshold individual component being representative of both the first threshold's variation, as well as any latent regression individual effects. Overall, controlling for unobservable factors thus equates to integrating out $J-1$ components. For individual i, the log-likelihood function is:

$$LnL_i = Ln \sum_{k=1}^{J} \int_{\omega_1,\ldots,\omega_{J-1}} \prod_{t=1}^{T} \left[\Phi(\mu_{ik} - x'_{it}\beta) - \Phi(\mu_{i,k-1} - x'_{it}\beta)\right]^{I(y=k)} \phi(\omega_1,\ldots,\omega_{J-1}) d\omega_1 \ldots d\omega_{J-1}$$

where μ_{i1} is as defined earlier, but incorporating the individual effect and constant from the latent regression. The simulated log-likelihood function thus takes the form:

$$LnL_s = \sum_{n=1}^{N} Ln \sum_{k=1}^{J} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \left[\Phi(\mu_{ikr} - x'_{it}\beta) - \Phi(\mu_{i,k-1,r} - x'_{it}\beta)\right]^{I(y=k)}$$

and is maximised, subject to β, and μ_{ikr} (the r^{th} draw for μ_{ik}) in order to get suitable estimates for these coefficients. The resulting specification is referred to as the individual thresholds specification, with thresholds treated as random. The assumed lack of correlation with latent regression covariates in this specification is imperative for allowing free interpretation of the effect of these covariates on the latent propensity. However, the model in its current form might be too inflexible to allow for an appropriate amount of variation in the thresholds, potentially affecting latent regression coefficient estimates in the process. In particular, factors that influence one threshold might be considered quite likely to influence each of the other thresholds. Thus drawing correlations between the thresholds is considered an appropriate way to incorporate additional heterogeneity, and better account for variation in survey respondent perceptions that may influence outcome reporting.

2.2 A Correlated Individual Thresholds Model

In order to allow for correlations between the individual components of the thresholds, a simulated maximum likelihood methodology is used, in similar vain to the methodology used by Greene (2007) to correlate random latent regression coefficients. The $(J-1)\times(J-1)$ variance-covariance matrix for the individual components, Ω, is decomposed as $\Omega = LD^2L'$, where L is lower triangular matrix with ones on the diagonal, and D is a diagonal matrix with strictly positive elements (Greene and Hensher, 2010). In order to calculate the likelihood function, we first calculate i.i.d normal draws for each of the the individual components for each individual, and pre-multiply by the appropriate function of the variance-covariance decomposition:

$$\omega_i^* = LDc_i$$

where c_i is an $J-1$ by R matrix of draws for individual i. Defining ω_{ijr}^* as the jth row and rth column of matrix ω_i^*, we then calculate the draws of each threshold as

$$\mu_{i1r}^* = \lambda_1 + \omega_{i1r}^*$$

$$\mu^*_{ijr} = \mu^*_{i,j-1} + e^{(\lambda_j + \omega^*_{ijr})}, \;\; 2 \le j \le J-1.$$

The implied log-likelihood is thus

$$LnL_i = Ln \sum_{k=1}^{J} \int_{\omega_1,\ldots,\omega_{J-1}} \prod_{t=1}^{T} \left[\Phi(\mu^*_{ik} - x'_{it}\beta) - \Phi(\mu^*_{i,k-1} - x'_{it}\beta)\right]^{I(y=k)} \phi(\omega_1,\ldots,\omega_{J-1}) d\omega_1 \ldots d\omega_{J-1}$$

Thus the simulated maximum likelihood function to be maximised is:

$$LnL_s = \sum_{n=1}^{N} Ln \sum_{k=1}^{J} \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \left[\Phi(\mu^*_{ikr} - x'_{it}\beta) - \Phi(\mu^*_{i,k-1,r} - x'_{it}\beta)\right]^{I(y=k)}$$

Partial effects of the latent regression variables can then be calculated as:

$$\frac{Pr(y_{it} = j | x_{it})}{dx_{it}} = \int_{\omega_1,\ldots,\omega_{J-1}} \left[\left[\phi(\mu^*_{ik} - x'_{it}\beta) - \phi(\mu^*_{i,k-1} - x'_{it}\beta)\right][-\beta]\right] \phi(\omega_1,\ldots,\omega_{J-1}) d\omega_1 \ldots d\omega_{J-1}$$

which accounts for the unobservable factors that influence the probabilities of each outcome, thus avoiding the issue of having to assume a value for the unobserved effect.

2.3 Extending to the Variance Function

As shown in applications, such as Ritter and Vance (2011), Lemp, Kockelman, and Unnikrishnan (2011) and Litchfield, Reilly, and Veneziani (2010), deterministic extensions from the standard Ordered Probit model to models that incorporate heteroscedasticity may be necessary to ensure consistency of the latent regression covariates. This is because there may be factors causing the underlying propensity for some individuals to be more extreme than in others. For example, individuals with personality disorders may exhibit more extreme health responses to external stimuli than mentally healthy individuals. As a result, we consider a parameterisation of the latent regression error variance, with the potential to include both a deterministic and stochastic (but non-time varying) component.

$$\sigma^2_{\epsilon_{it}} = [e^{(w'_i\theta + \omega_J)}]^2$$

w_i cannot include a constant in order to maintain identification between the variance and latent regression coefficients. Along with individual thresholds, such a specification would imply individual log-likelihoods of:

$$LnL_i = Ln\sum_{k=1}^{J}\int_{\omega_1,\ldots,\omega_J}\prod_{t=1}^{T}\left[\Phi\left(\frac{\mu_{ik}^* - x_i'\beta}{e^{(w_{it}'\theta+\omega_{iJ})}}\right) - \Phi\left(\frac{\mu_{i,k-1}^* - x_i'\beta}{e^{(w_{it}'\theta+\omega_{iJ})}}\right)\right]^{I(y=k)}\phi(\omega_1,\ldots,\omega_J)d\omega_1\ldots d\omega_J$$

Implying the simulated log-likelihood function:

$$LnL_s = \sum_{n=1}^{N} Ln\sum_{k=1}^{J}\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\left[\Phi\left(\frac{\mu_{ikr}^* - x_i'\beta}{e^{(w_{it}'\theta+\omega_{iJr})}}\right) - \Phi\left(\frac{\mu_{i,k-1,r}^* - x_i'\beta}{e^{(w_{it}'\theta+\omega_{iJr})}}\right)\right]^{I(y=k)}$$

where μ_{ikr}^* are defined as in section 2.2 and replacing ω_i^* with:

$$\omega_i^* = LDc_i$$

where c_i is now a J by R matrix of draws for individual i (an additional column of draws).

This specification nests the previously specified stochastic threshold models, a model with stochastic variance and deterministic thresholds, and also allows for correlations to be drawn between the thresholds and the variance. We motivate these correlations on two bases. Firstly, there may be underlying factors that influence both how individuals categorise their health, as well as how sensitive their level of underlying health is to external factors. The second motivation pertains to the nature of any unobserved thresholds covariates in the HOPIT specification, and the general flexibility allowed by correlated stochastic threshold models. As the HOPIT model is ordinal in nature, an increase in the variance of the latent regression might likewise be interpreted as a decrease in the scale of the thresholds. Correlations between the variance and thresholds might, therefore simply account for shared variation between the thresholds. Hence, correlating the threshold components might have the effect of allowing these components to be more free to account for idiosyncratic variation, as opposed to accounting for variation that is inherently shared.

3 Monte Carlo Simulations

Although it is known that simulated maximum likelihood obtains consistent parameter estimates asymptotically, the issue of unobservable threshold covariates is more likely to be prominent in smaller datasets. While assumptions might be able to be made about suitable covariates in datasets such as the British Household Panel Survey (BHPS) and Household, Income and Labour Dynamics Survey of Australia (HILDA), defensible covariates are likely to be more difficult to find in smaller surveys and cohort studies. Even in larger datasets, it would be useful to know how the simulated maximum likelihood methodology would behave if utilised on a small sub-sample of interest. One of the primary benefits of the maximum likelihood estimation methodology is the asymptotic performance of parameter estimates. However, these simulations intend to provide insight into how to evaluate the performance of stochastic threshold models in the finite sample case.

In addition, as many assumptions are required in the specification of such models, such as the form of the unobserved variables, and the functional form of the thresholds, Monte Carlo simulations are run to assess model performance under misspecification.

In all simulations, a constant variance is assumed, with the focus being on the uncorrelated and correlated individual threshold models defined in sections 2.1 and 2.2. While simulations could have been extended to specifications with individual variances, we do not do so for two reasons. Firstly, the key objective of the simulations is to find what aspects of estimation results are likely to be indicative of a good model estimated using the techniques we employ. An extension to individual variances is considered unnecessary. Secondly, the issue of unobserved threshold covariates is considered more pertinent in the literature, due to the threshold covariates often being related to respondent perceptions, while variance covariates are more reasonably motivated along the same lines as threshold covariates.

3.1 The Data Generating Process

Data is assumed to be generated along the lines of the ordered choice framework with thresholds varying from distinct covariates that are uncorrelated to the latent regression variables and error term. A three-outcome dependent variable is assumed for simplicity. All components of the simulation ($x_1, x_2^*, \epsilon, z_1, z_2$ and α) are generated via a multivariate normal distribution in order to allow more control over correlations between model components. They also motivated under the previously discussed assumption that the empirical cumulation of factors affecting each of thresholds approximating a normal distribution. The mean of each variable is equal to zero. In addition, the covariance matrix is constructed as:

$$\Sigma = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2^*} & \sigma_{x_1\epsilon} & \sigma_{x_1z_1} & \sigma_{x_1z_2} & \sigma_{x_1\alpha} \\ \sigma_{x_2^*x_1} & \sigma^2_{x_2^*} & \sigma_{x_2^*\epsilon} & \sigma_{x_2^*z_1} & \sigma_{x_2^*z_2} & \sigma_{x_2^*\alpha} \\ \sigma_{\epsilon x_1} & \sigma_{\epsilon x_2^*} & \sigma^2_{\epsilon} & \sigma_{\epsilon z_1} & \sigma_{\epsilon z_2} & \sigma_{\epsilon\alpha} \\ \sigma_{z_1x_1} & \sigma_{z_1x_2^*} & \sigma_{z_1\epsilon} & \sigma^2_{z_1} & \sigma_{z_1z_2} & \sigma_{z_1\alpha} \\ \sigma_{z_2x_1} & \sigma_{z_2x_2^*} & \sigma_{z_2\epsilon} & \sigma_{z_2z_1} & \sigma^2_{z_2} & \sigma_{z_2\alpha} \\ \sigma_{\alpha x_1} & \sigma_{\alpha x_2^*} & \sigma_{\alpha\epsilon} & \sigma_{\alpha z_1} & \sigma_{\alpha z_2} & \sigma^2_{\alpha} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \sigma_{z_1z_2} & 0 \\ 0 & 0 & 0 & \sigma_{z_1z_2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\alpha} = 2.2 \end{bmatrix}$$

to isolate the effect of the true threshold correlation, $\sigma z_1 z_2$. The variance of α is chosen to reflect typical estimation results from empirical work; in particular, preliminary estimations of models on HILDA, which is used for application later in this research. We consider a panel data case in simulations, to allow the results to be more relevant to empirical applications. T time periods, with N observations per time period are specified in each group of simulations. The observations for z_{1it}, z_{2it} and α are restricted to be the same across all time periods. The dependent variable is thus generated according to the process:

$$y^*_{it} = \beta_0 + x_{1it}\beta_1 + x_{2it}\beta_2 + \alpha_i + \epsilon_{it}$$

where $x_{2it} = 1(x^*_{2it} > 0)$. Thus, for the dependent variable,

$$y_{it} = \left\{ \begin{array}{ll} 0 & \text{if } y^*_{it} < \mu_{i1} \\ 1 & \text{if } \mu_{i1} \leq y^*_{it} < \mu_{i2} \\ 2 & \text{if } \mu_{i2} \leq y^*_{it} \end{array} \right\}$$

and

$$\mu_{i1} = z_{1i}\delta_1,$$

$$\mu_{i2} = \mu_{i1} + e^{(\delta_2 + z_{2i}\delta_3)}.$$

Values for the vector of coefficients are $\{\beta_0, \beta_1, \beta_2, \delta_1, \delta_2, \delta_3\}' = \{0.5, -0.2, 0.7, 0.5, 0.4, -0.8\}'$, and are chosen to construct an unconditional distribution of y^*_{it} that is well defined (that is, at least one observation of each outcome in most generated samples). Samples not satisfying this basic condition are re-drawn.

3.1.1 Model Misspecification

In order to assess the sensitivity of the models to specific forms of misspecification, simulations are run for two additional scenarios - threshold covariate distribution misspecification, and threshold functional form misspecification.

Covariate Distribution Misspecification

A motivation for assuming normally distributed individual components is that the cumulation of unobservable factors that would realistically influence the thresholds would converge to normality, at least asymptotically. A useful question to ask, therefore, is what if this is not the case? To test this, bi-modal threshold component distributions are considered. These combine a normally distributed variable, as well as a dichotomous variable. Thus the thresholds are of the form:

$$\mu_{i1} = z_{1i}\delta_i,$$

$$\mu_{i2} = \begin{cases} \mu_{i1} + e^{\delta_2 + z_{i2}\delta_3 + \delta_4} & \textit{with probability } 0.5 \\ \mu_{i1} + e^{\delta_2 + z_{i2}\delta_3} & \textit{with probability } 0.5 \end{cases}$$

with δ_4 set to the value -0.5, leaving the overall distribution of $z_{i2}\delta_3 + \delta_4$ to be bimodal, yet somewhat close to normal.

Threshold Functional Form Misspecification

Secondly, a linear thresholds case is considered to look at the sensitivity of the modeling framework to threshold structures that are not necessarily in HOPIT form. That is, thresholds are specified along the lines of Pudney and Shields (2000):

$$\mu_{i1} = z_{1i}\delta_1,$$

$$\mu_{i2} = \delta_2 + z_{2i}\delta_3$$

Of course, with the models estimated taking a HOPIT form in the thresholds, this also implies threshold distribution misspecification. With linear thresholds there is the possibility of thresholds crossing over, which would lead to negative probabilities. Thus, at each replication, probabilities are checked and, if any negative probabilities are detected, the sample is re-drawn. Coefficient values are chosen to avoid this occurring, and in practice less than 0.5% of samples contained negative probabilities.

3.2 Performance Measures

For each of the replications, the standard Ordered Probit model, the HOPIT model with true threshold covariates, the uncorrelated individual thresholds HOPIT model (referred to as model UIT) and the correlated individual thresholds HOPIT model (model CIT) are estimated. Multiple measures of the fit of each of these models are considered in order to understand the dynamics underlying each model.

3.2.1 Coefficient Mean-Squared Error

The latent regression parameters are of key interest to analysts in most studies. Thus the mean-squared error (MSE) of each of the βs is calculated for each set of parameters. For this study, the MSEs are calculated using as:

$$\widehat{MSE_{\beta_l,model}} = \frac{1}{S}\sum_{h=1}^{S}\left[\widehat{\beta_{l,h,m}} - \widehat{\beta_{l,h,HOPIT}}\right], \quad l = 1, 2$$

where S is the total number of replications (for this study we set $S = 1000$), and $\widehat{\beta_{l,h,m}}$ is the coefficient of β_l estimated in replication h of model $m = \{OP, uncorrindthres, corrindthres\}$. For mean squared error calculations, the "error" considered is the difference between the parameter estimate under the alternative model, m, and the parameter estimate of a HOPIT model, in which the true covariates have been utilised, $\widehat{\beta_{l,h,HOPIT}}$. As the HOPIT model with true threshold covariates is the best model that would be estimable by the analyst, the MSE thus signifies the closeness of the other models' parameter estimates with the best possible outcome, and negates the effect of sample variation that would otherwise cloud the results shown with differing sample sizes and orders of integration. As the MSE for the HOPIT model with true threshold covariates will be zero by construction, these are not reported.

Table 1: Monte Carlo Model Parameter Summary

Parameter Estimate	**Description**
$\widehat{\beta_{l,h,HOPIT}}$	Estimate from Hopit model with true covariates
$\widehat{\beta_{l,h,OP}}$	Estimate from Ordered Probit model
$\widehat{\beta_{l,h,UIT}}$	Estimate from Uncorrelated Individual Threshold Model
$\widehat{\beta_{l,h,CIT}}$	Estimate from Correlated Individual Threshold Model

3.2.2 Log-Likelihoods and Likelihood-Ratio Tests

Taking advantage of the inherent nesting of some of the models, likelihood ratio (LR) tests can be used to determine whether individual thresholds are preferred over non-varying thresholds, and whether correlations between the individual thresholds are merited. Thus likelihoods are recorded for each model estimation, so that likelihood ratio tests can used to assess the suitability of threshold components.

3.3 Results

The Monte Carlo results have some important implications for health economists. In terms of overall fit, in all sets of simulations, accounting for unobserved heterogeneity in the thresholds proved significantly better than assuming the standard Ordered Probit every time. This is shown by the "% rejected" statistic in each of the results tables, and implies that, without access to appropriate threshold variables, analysts should attempt at least one type of model incorporating unobserved heterogeneity.

The assessment of correlations across threshold components is much more difficult, as likelihood ratio tests exhibit poor power at these sample sizes. This is reflected with the "% rejected" statistic sitting around 10% to 30% in most simulations, regardless of the true data generating process incorporating non-zero correlations. The best explanation the authors can offer is that sample sizes are not yet large enough for the asymptotics to be valid for likelihood ratio tests. This may be exacerbated by the collection of the latent regression individual component (which is uncorrelated to the thresholds, with the first threshold component, diluting the effect of this component. In addition, it is possible that the uncorrelated threshold model systematically over-fits the data when sample sizes are small. Several factors improved the correct rejection of the false null hypotheses, including increasing the sample size, increasing the degree of integration (number of Halton draws), and increasing the underlying (true) correlation between the components. The result summaries not shown here can be found in Appendix 3.

In practice, this under-rejection of the null hypothesis implies two things. Firstly, if an estimated model does result in significant correlation coefficients, the analyst should be reasonably confident that true correlations are evident. Secondly, it is suggested that if the null hypothesis is not rejected, information criteria should be used as an alternative measure of model fit. For example, with simulation set 3, for correlations of 0.3 and 0.7, although the LR test resulted in rejection 17% and 30% of the time, the Alkaike Information Criteria (AIC) suggested the use of correlated threshold components

Table 2: Set 3 - N = 1000, M = 100, T = 8

Measure	**Model**	$\sigma_{\mathbf{Z_1Z_2}} = \mathbf{0}$	**0.3**	**0.7**	**-0.7**
β_1 **MSE**	OP	1.9E-05	1.2E-05	1.8E-06	1.3E-05
	UIT	1.2E-06	2.0E-06	3.5E-06	2.7E-06
	CIT	2.6E-06	2.6E-06	2.7E-06	2.2E-06
β_2 **MSE**	OP	2.3E-04	1.4E-04	1.8E-05	1.6E-04
	UIT	1.2E-05	2.3E-05	4.1E-05	3.3E-05
	CIT	3.0E-05	3.0E-05	3.1E-05	2.6E-05
Log-Likelihood	True Model	-4979.3	-5032.3	-5214.7	-5082.2
	OP	-5591.2	-5618.6	-5600.0	-5453.1
	UIT	-5356.7	-5393.9	-5483.5	-5360.5
	CIT	-5356.0	-5392.9	-5481.9	-5359.4
LR Test (% Rejected)	UIT vs OP	100%	100%	100%	100%
	CIT vs UIT	12%	17%	30%	21%

Table 3: Misspecification - N = 1000, M = 100, T = 8

Measure	**Model**	**Bi-modal Normal**	**Linear** $\sigma_{\mathbf{Z_1Z_2}} = 0$	**Linear** 0.3	**Linear** 0.7
β_1 **MSE**	OP	4.5E-06	1.1E-06	1.4E-06	2.6E-06
	UIT	4.7E-06	5.7E-06	8.1E-06	1.4E-05
	CIT	2.8E-06	2.7E-06	3.1E-06	3.2E-06
β_2 **MSE**	OP	5.2E-05	1.2E-05	1.5E-05	3.1E-05
	UIT	5.6E-05	7.2E-05	9.9E-05	1.7E-04
	CIT	3.3E-05	3.4E-05	3.6E-05	4.1E-05
Log-Likelihood	True Model	-5138.9	-5343.2	-5231.5	-5007.7
	OP	-5601.4	-5590.0	-5608.6	-5625.4
	UIT	-5446.9	-5553.7	-5556.9	-5551.2
	CIT	-5445.2	-5552.0	-5554.7	-5547.5
LR Test (% Rejected)	UIT vs OP	100%	100%	100%	100%
	CIT vs UIT	32%	33%	43%	68%

49% and 65% of the time respectively.

Results also give some indication about appropriate strategies for obtaining accurate latent regression variable estimates. For most simulation sets, Ordered Probit models appear to out-perform the unobserved heterogeneity models when correlations are strong and positive. This is likely due to the joint variation between the threshold and latent regression effect being absorbed by the single individual effect in the Ordered Probit model. In each of the other cases, however, models incorporating unobserved heterogeneity result in better latent regression coefficient estimates. Uncorrelated thresholds generally tend to result in lower levels of MSE than for correlated thresholds. This is a result of limited sample sizes affecting the estimation of these models. Consequently, simulation table 2 shows that when the sample size is increased, and the correlation increased to 0.7, the MSE for the correlated threshold model falls below that of the uncorrelated model. Overall, results indicate that, for improving the accuracy of latent regression parameters, the uncorrelated thresholds model seems to work better than its' correlated counterpart. However, if the objective is to improve overall model fit, or incorporate heterogeneity more generally, then the correlated model should at least be tried. It is at this juncture that the authors reiterate that these are small sample results. When larger datasets are available, we suggest relying on the asymptotic properties of the models to justify the inclusion of correlated thresholds over constraining the model to uncorrelated thresholds.

Simulations for misspecification indicate that, if it is suspected that the true threshold functional form is linear, standard Probit models tend to result in better latent regression coefficient estimates than both the unobserved threshold specifications. However, the unobserved threshold models fit the data much better, and correlating the thresholds is justified across a range of true correlation values. Even in this case, the poor power of the LR test suggests that coefficient significance should not be the sole indicator of model performance.

Under the bi-modal normal distribution of the covariates, the correlated thresholds model yielded a better fit for the latent regression parameters, with an MSE falling by more than 40%. The same overall model fit issues arise as in the other simulations, however, with the LR test rejecting the null hypothesis 32% of the time, when the true correlation between the continuous components was 0.7. In conjunction with the functional form results, this indicates that when the analyst may be unsure of the true data generating process (a majority of cases, as linearity is often assumed for simplicity, rather than motivated on some economic basis), correlating unobserved components is an effective way to allow the thresholds to fit the data better.

4 Application to Self-Assessed Health

An important application is now considered; specifically, that of assessing the determinants of underlying health of survey respondents. In many cases a Self-Assessed Health (SAH) variable is utilised. Although other measures have been considered within the literature (for example, Ziebarth (2011) and Lecluyse and Cleemput (2006)), SAH has been shown to be an accurate indicator of underlying health (Idler and Benyamini, 1997; Benyamini, Leventhal, and Leventhal, 1999). Of course, one of the key criticisms of SAH as a subjective measure of underlying health is the scope for subjective influences to mask the true level of underlying health of respondents. Several models for SAH are thus estimated, using waves one to eight of the HILDA. A baseline model for underlying health is assumed, along the lines of Hernndez-Quevedo, Jones, and Rice (2008), who motivate (broadly) the use of marital status, ethnicity, education, household dynamics, income and age as covariates for health. Variable descriptions are given in Appendix 1. Although other variables could be included, such as other health or unemployment variables (Powdthavee, 2009), they are avoided primarily for reasons of parsimony and endogeneity (Weterings et al., 2011).

As mentioned, SAH provides large scope for the influence of threshold covariates due to the subjective nature of the variable. If vignette questions were incorporated into the HILDA dataset, such techniques might be used for clarifying the true effect of each of the covariates on underlying health. As with most large datasets, this is not the case. It might also be preferable to directly include variables that fulfill the criteria of influencing the thresholds, but not influencing underlying health. As argued by Weterings et al. (2011), such variables might include factors such as personality. However, while indicators for personality (the 'big five' personality factors Losonscz (2004)) are collected in HILDA, they are not measured in every wave of the data. As the traits are measured via the use of a self-evaluation questionnaire Losonscz (2004), it is possible that measurement error could arise from inaccurate perceptions of individual's own traits. If the traits were measured each year an assumption of randomness in variable measurement error might suffice for empirical purposes. However, if the same variable values were used each year, systematic biases would be constructed. The motivation of other possible variables are discussed in detail within Weterings et al. (2011). The conclusion is reached that, within the panel setting, such variables are likely to be inappropriate in practice.

Individual component models are considered necessary in order to account for differences in respondent perceptions of what constitutes different categories of health. Models that incorporate time invariant, individual-specific thresholds are estimated to control for factors such as personality, and better identify these effects. In addition, drawing correlations between the thresholds might be necessary in order to best account for joint variation in thresholds in response to these unobservable threshold factors, and to ease any potential specification issues related to the non-linear form of the thresholds, as suggested by the Monte-Carlo study in section 3. After cleaning the data of missing observations, the working dataset consisted of an unbalanced panel of 18667 individuals over 8 waves of data.

5 Estimation Results

Overall, six different stochastic component models are estimated. These models are explained in table 4, with the REOP indicating a standard Random Effects Ordered Probit, and the other specifications being variants of the individual thresholds and individual variance models suggested in this paper.

Table 4: Model Name Summary

Model Name	**Individual thresholds**	**Individual Variance**	**Thresholds Correlated**	**Thresholds/Variance Correlated**
REOP	No	No	No	-
IC1	Yes	No	No	-
IC2	Yes	No	Yes	-
IC3	No	Yes	-	-
IC4	Yes	Yes	No	No
IC5	Yes	Yes	Yes	No
IC6	Yes	Yes	Yes	Yes

The models are assessed via an examination of coefficient direction and significance, as well as comparison of likelihood functions and likelihood ratio tests, where appropriate. While the simulation results suggested poor power of LR ratio tests, the large sample size used for estimation of these models is likely to make the asymptotic maximum likelihood properties more relevant. In addition, if poor power is still exhibited with large sample sizes, and a test indicates joint significance of an extension, the analyst should be even more confident of the result. It should be noted that model IC3 incorporates individual effects in the latent regression function, although these might foreseeably account for factors that influence the first threshold. This is done to allow better comparisons with each of the other models, which also account for the latent regression individual effect. These effects are assumed to be uncorrelated to the individual variance for model IC3.

5.1 Latent Regression Coefficients

Latent regression coefficients provide subtle indications that model choice is important when it comes to modeling SAH. While being married is always found to be associated with a statistically significant increase in underlying health, the effect of being separated varies across models. For example, with the REOP specification, being separated is not found to have a significant effect at the 5% level. However, once individual thresholds are accounted for, such an effect is found, and with threshold correlations allowed the effect is significant at the 1% level. This suggests the existence of factors that influence some separated individuals to report their level of health differently to other individuals, even when the underlying level of health is the same.

A similar, albeit more extreme, finding is made for the health levels of immigrants. While the REOP and uncorrelated individual thresholds models suggest no difference between the health levels

Table 5: Latent Regression Parameter Estimates

Variable	REOP	IT1	IT2	IT3	IT4	IT5	IT6
Constant	5.073**	5.049**	5.957**	5.901**	5.855**	5.961**	6.133**
	(0.068)	*(0.137)*	*(0.088)*	*(0.084)*	*(0.083)*	*(0.089)*	*(0.100)*
Married	0.090**	0.096**	0.084**	0.085**	0.081**	0.083**	0.086**
	(0.023)	*(0.029)*	*(0.025)*	*(0.027)*	*(0.027)*	*(0.025)*	*(0.026)*
Separated	0.063	0.096*	0.121**	0.109**	0.106**	0.120**	0.128**
	(0.032)	*(0.038)*	*(0.035)*	*(0.038)*	*(0.038)*	*(0.035)*	*(0.036)*
Immigrant	0.037	0.008	0.067*	0.050	0.036	0.066*	0.091**
	(0.028)	*(0.042)*	*(0.031)*	*(0.035)*	*(0.033)*	*(0.031)*	*(0.031)*
Postgraduate	0.486**	0.501**	0.537**	0.570**	0.591**	0.536**	0.511**
	(0.065)	*(0.063)*	*(0.062)*	*(0.066)*	*(0.067)*	*(0.062)*	*(0.068)*
Bachelors	0.377**	0.366**	0.353**	0.435**	0.429**	0.354**	0.356**
	(0.032)	*(0.036)*	*(0.034)*	*(0.037)*	*(0.037)*	*(0.035)*	*(0.035)*
Other Tertiary	0.089**	0.072**	0.069**	0.106**	0.100**	0.070**	0.080**
	(0.023)	*(0.027)*	*(0.025)*	*(0.027)*	*(0.027)*	*(0.025)*	*(0.025)*
Year 12	0.061*	0.043	0.026	0.073*	0.079**	0.027	0.023
	(0.025)	*(0.027)*	*(0.027)*	*(0.030)*	*(0.029)*	*(0.027)*	*(0.027)*
Household Size	−0.017*	−0.018	−0.018*	−0.018*	−0.018*	−0.018*	−0.017*
	(0.007)	*(0.013)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*
Children 0-4	−0.055**	−0.052*	−0.052*	−0.049*	−0.056**	−0.053**	−0.056**
	(0.019)	*(0.023)*	*(0.020)*	*(0.022)*	*(0.021)*	*(0.020)*	*(0.020)*
Children 5-9	0.022	0.028	0.019	0.031	0.028	0.018	0.015
	(0.018)	*(0.022)*	*(0.019)*	*(0.020)*	*(0.020)*	*(0.019)*	*(0.019)*
Children 10-14	0.081**	0.085**	0.090**	0.093**	0.092**	0.090**	0.086**
	(0.017)	*(0.022)*	*(0.019)*	*(0.019)*	*(0.019)*	*(0.019)*	*(0.019)*
Ln(Income)	0.126**	0.123**	0.110**	0.124**	0.124**	0.111**	0.109**
	(0.011)	*(0.017)*	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*
Ln(Income)^2	0.010**	0.009**	0.008**	0.009**	0.010**	0.008**	0.008**
	(0.001)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*
Age	−0.742**	−0.785**	−0.677**	−0.836**	−0.812**	−0.676**	−0.715**
	(0.013)	*(0.114)*	*(0.065)*	*(0.071)*	*(0.068)*	*(0.065)*	*(0.065)*
Age^2	0.000	0.007	−0.031*	−0.007	−0.010	−0.031*	−0.024
	(0.001)	*(0.023)*	*(0.013)*	*(0.015)*	*(0.014)*	*(0.013)*	*(0.013)*
Log-Likelihood	-96431.8	-94969.7	-94256.0	-94384.0	-94379.7	-94228.0	-94204.9

** *Significant at 1% significance level*   * *Significant at 5% significance level*

Standard errors in parentheses

Table 6: Other Parameter Estimates

Parameter	REOP	IT1	IT2	IT3	IT4	IT5	IT6
μ_2	1.810**	0.441**	0.790**	0.764**	0.758**	0.792**	0.891**
	(0.067)	*(0.015)*	*(0.021)*	*(0.011)*	*(0.011)*	*(0.021)*	*(0.027)*
μ_3	3.729**	0.630**	0.691**	0.791**	0.782**	0.692**	0.714**
	(0.066)	*(0.007)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.010)*
μ_4	5.749**	0.779**	0.692**	0.853**	0.851**	0.693**	0.669**
	(0.066)	*(0.007)*	*(0.010)*	*(0.008)*	*(0.008)*	*(0.010)*	*(0.011)*
$\mathbf{D_{11}}$	1.407**	1.424**	2.136**	1.757**	1.748**	2.134**	2.163**
	(0.019)	*(0.015)*	*(0.039)*	*(0.016)*	*(0.016)*	*(0.039)*	*(0.045)*
$\mathbf{D_{22}}$		0.376**	0.358**		0.016**	0.356**	0.237**
		(0.012)	*(0.010)*		*(0.020)*	*(0.010)*	*(0.014)*
$\mathbf{D_{33}}$		0.379**	0.259**		0.093**	0.257**	0.290**
		(0.009)	*(0.017)*		*(0.017)*	*(0.017)*	*(0.021)*
$\mathbf{D_{44}}$		0.383**	0.023**		0.108**	0.024**	0.071**
		(0.011)	*(0.021)*		*(0.021)*	*(0.020)*	*(0.027)*
$\mathbf{D_{55}}$				0.409**	0.402**	0.037**	0.023**
				(0.006)	*(0.007)*	*(0.029)*	*(0.024)*
$\mathbf{L_{21}}$			0.119**			0.119**	0.084**
			(0.004)			*(0.005)*	*(0.005)*
$\mathbf{L_{31}}$			0.130**			0.130**	0.083**
			(0.006)			*(0.006)*	*(0.006)*
$\mathbf{L_{32}}$			0.431**			0.436**	0.292**
			(0.050)			*(0.050)*	*(0.110)*
$\mathbf{L_{41}}$			0.107**			0.108**	0.122**
			(0.008)			*(0.008)*	*(0.011)*
$\mathbf{L_{42}}$			−0.422**			−0.410**	−0.992**
			(0.056)			*(0.059)*	*(0.110)*
$\mathbf{L_{43}}$			1.191**			1.198**	1.119**
			(0.100)			*(0.101)*	*(0.123)*
$\mathbf{L_{51}}$							−0.022**
							(0.006)
$\mathbf{L_{52}}$							−0.825**
							(0.118)
$\mathbf{L_{53}}$							0.219**
							(0.086)
$\mathbf{L_{54}}$							−0.236**
							(0.346)

** *Significant at 1% significance level*   * *Significant at 5% significance level*

Standard errors in parentheses

of immigrants and Australian-born respondents, allowing correlations between thresholds contradicts the findings the REOP and IC1 models would have made, had they been the final models used. The additional flexibility allowed by correlating variances to threshold components provides further evidence of an effect, with significance at the 1% level. As a result, the choice of model is seen to be very important in the investigation of health determinants.

Other latent regression covariates show logical signs and significance. Education has a positive and increasing association with underlying health, while increases in household size are shown to have a negative effect; Children under five years old are associated with poorer health, while children over ten years old are associated with higher levels of health; and income and age have the expected positive and negative relationships with health respectively. The key output from these results, however, is that choice of model does have an effect on the significance of some variables.

5.2 Unobserved Heterogeneity Coefficients and Overall Model Fit

Corresponding to the changes across the latent regression coefficients, the unobserved heterogeneity components show strong significance in most cases, and improvements in the likelihood function are evident for each extension. An interesting result is the lack of significance of the variance component for models IC5 and IC6, regardless of the significant correlations between the variance component and the threshold components in model IC6. A likelihood ratio test suggests strong joint significance of the threshold-variance correlation parameters. Models IC2 and IC5 (correlations between thresholds) are also strongly preferred over the uncorrelated models, with all unobserved component models strongly preferred over a simple REOP model (minimum LR statistic = 2924 with 4 degrees of freedom).

5.3 Post-Estimation Analysis

Supplementing coefficient interpretation, several post-estimation analysis techniques are employed to get further insight into where the additional flexibility allowed by the more advanced models, is utilised by the data. In order to do so, marginal effects and probability plots (demonstrating the effect of model choice on prediction of outcome probabilities across different ages) are constructed. In particular, the transitions from models IC4 to IC5 and models IC5 to IC6 are assessed, in order to get a sense of the effect of the two key extensions offered by this paper.

5.3.1 Marginal Effects

In order to investigate changes in the effects of latent regression variables on each of the health categories, marginal effects are calculated at sample means. In order to avoid having to assume particular values for the unobserved effects (and as with the hit-miss tables), unobserved heterogeneity is integrated out in the calculation of the marginal effects. Standard errors are calculated using the delta method.

Table 7: Marginal Effects at Sample Means: Model IC6

Variable	Poor	Fair	Good	Very Good	Excellent
Married	-0.21	-0.72	-0.72	0.84	0.81
	(0.07)	*(0.23)*	*(0.23)*	*(0.27)*	*(0.26)*
Separate	-0.32	-1.06	-1.06	1.24	1.19
	(0.10)	*(0.32)*	*(0.32)*	*(0.38)*	*(0.36)*
Immigran	-0.20	-0.68	-0.67	0.79	0.76
	(0.09)	*(0.29)*	*(0.29)*	*(0.34)*	*(0.32)*
Postgrad	-1.47	-4.94	-4.93	5.78	5.56
	(0.18)	*(0.58)*	*(0.60)*	*(0.70)*	*(0.65)*
Bachelor	-0.97	-3.25	-3.24	3.80	3.66
	(0.10)	*(0.33)*	*(0.33)*	*(0.40)*	*(0.36)*
Otherte	-0.20	-0.68	-0.68	0.80	0.77
	(0.07)	*(0.24)*	*(0.24)*	*(0.28)*	*(0.26)*
Educyr1	-0.11	-0.36	-0.36	0.42	0.40
	(0.07)	*(0.25)*	*(0.25)*	*(0.29)*	*(0.28)*
HHsize	0.04	0.14	0.14	-0.17	-0.16
	(0.02)	*(0.07)*	*(0.07)*	*(0.08)*	*(0.08)*
children	0.15	0.52	0.52	-0.60	-0.58
	(0.06)	*(0.19)*	*(0.19)*	*(0.22)*	*(0.21)*
children	-0.05	-0.16	-0.16	0.18	0.18
	(0.05)	*(0.18)*	*(0.18)*	*(0.20)*	*(0.19)*
children	-0.23	-0.78	-0.77	0.91	0.87
	(0.05)	*(0.17)*	*(0.17)*	*(0.20)*	*(0.19)*
lninc	-0.27	-0.89	-0.89	1.04	1.00
	(0.03)	*(0.10)*	*(0.10)*	*(0.12)*	*(0.10)*
age	2.22	7.44	7.41	-8.70	-8.38
	(0.08)	*(0.19)*	*(0.23)*	*(0.26)*	*(0.19)*

Note: Standard errors in parentheses

Table 7 details marginal effects with standard errors for model IC6, while other marginal effects can be obtained by contacting the authors. Marginal effects look similar across each of the models. However, the similarity in marginal effect magnitude is more reflective of the scale of the variables. In order to get a better idea of changes in estimated marginal effects, ratios of the effects between models are constructed. Tables 8 and 9 show the ratios of marginal effects between models IC5 and IC4,

Table 8: Marginal Effect Ratios - Models IC5/IC4

	Poor	**Fair**	**Good**	**Very Good**	**Excellent**
Married	99.48	75.05	98.22	79.43	96.06
Separated	117.78	88.85	116.28	94.04	113.72
Immigrant	181.93	137.24	179.61	145.24	175.66
Postgraduate	115.25	86.94	113.78	92.01	111.28
Bachelors	108.98	82.21	107.59	87.01	105.22
Other Tertiary	86.62	65.34	85.51	69.15	83.63
Year 12	68.98	52.04	68.11	55.08	66.61
Household size	104.08	78.52	102.75	83.10	100.49
Children 0-4	114.68	86.52	113.22	91.56	110.73
Children 5-9	93.16	70.28	91.96	74.37	89.95
Children 10-14	117.80	88.87	116.30	94.05	113.74
Ln(Income)	111.16	83.86	109.74	88.75	107.33
Age	115.59	87.20	114.12	92.28	111.60

and the ratios of marginal effects between models IC6 and IC5. For example, the effect of being an immigrant (with otherwise average characteristic) on the probability of selecting *Good* health would be 79.61

Table 9: Marginal Effect Ratios - Models IC6/IC5

	Poor	**Fair**	**Good**	**Very Good**	**Excellent**
Married	103.45	112.88	113.53	111.84	111.83
Separated	99.08	108.11	108.74	107.12	107.11
Immigrant	122.58	133.76	134.53	132.53	132.52
Postgraduate	91.55	99.90	100.47	98.98	98.97
Bachelors	88.14	96.18	96.73	95.29	95.29
Other Tertiary	89.11	97.24	97.80	96.34	96.34
Year 12	102.97	112.36	113.01	111.33	111.32
Household size	87.67	95.66	96.22	94.78	94.78
Children 0-4	106.02	115.69	116.36	114.63	114.62
Children 5-9	80.21	87.52	88.04	86.72	86.71
Children 10-14	89.78	97.97	98.53	97.07	97.06
Ln(Income)	89.98	98.18	98.75	97.28	97.27
Age	91.88	100.26	100.84	99.34	99.33

There are two important patterns recognised in Table **??**. The first regards changes in the magnitude of marginal effects across all of the outcome categories; a result of changes in estimated coefficients, and the most obvious outcome of increased heterogeneity allowed through threshold variation.

The second pattern regards the magnitude of marginal effects within each of the outcome categories, but across all of the covariates. Namely, in the transition from model IC4 to IC5, marginal effects ratios for each of the covariates on the *Poor*, *Good* and *Excellent* outcomes are larger than for the *Fair* and *Very Good* health outcomes. Thus, allowing correlations across threshold components allows complex structural changes in marginal effects. With the transition from model IC5 to IC6, marginal effect ratios are more uniform across outcome categories, with the exception of the *Poor* health category, for which marginal effect ratios are lower.

5.3.2 Probability Plots

We also construct probability plots to show the changes in outcome probabilities across different ages. Other covariates are set at sample averages and individual components are once again integrated out, so that no assumptions about the values of unobserved covariates are made. Figure 1 shows estimated probabilities for each outcome across ages for model IT6. Other plots can be obtained by contacting the authors.

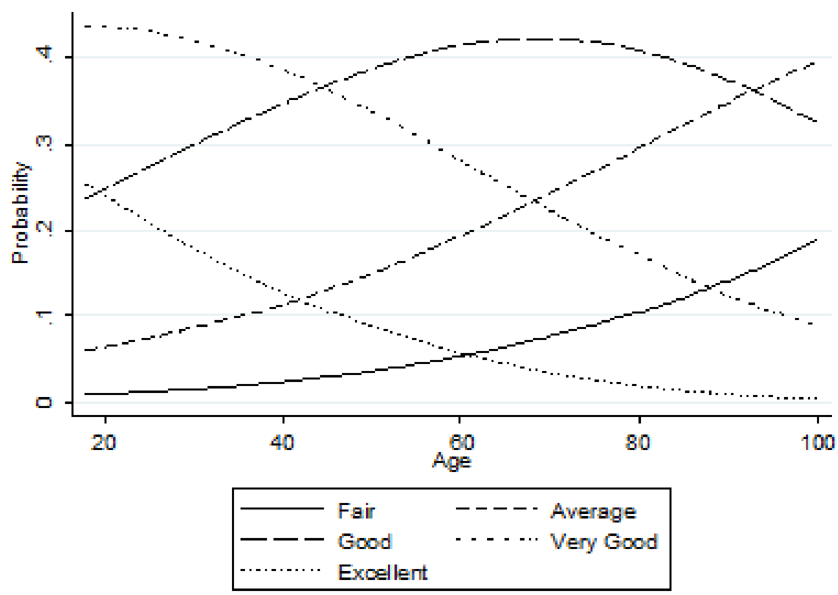

Figure 1: Probability of Health Outcomes for Model IT6

As probability plots appear similar across models, probability ratio plots are also shown, where deviations from zero indicate a difference in the estimated probability of the outcome category between models. From figure 2 it is seen that, for all ages, an individual with average characteristics would be predicted to be more likely to choose *poor* health if thresholds are allowed to be correlated. In addition, the probability of choosing *fair* health would be smaller, and *excellent* health would be much less likely for individuals over 50.

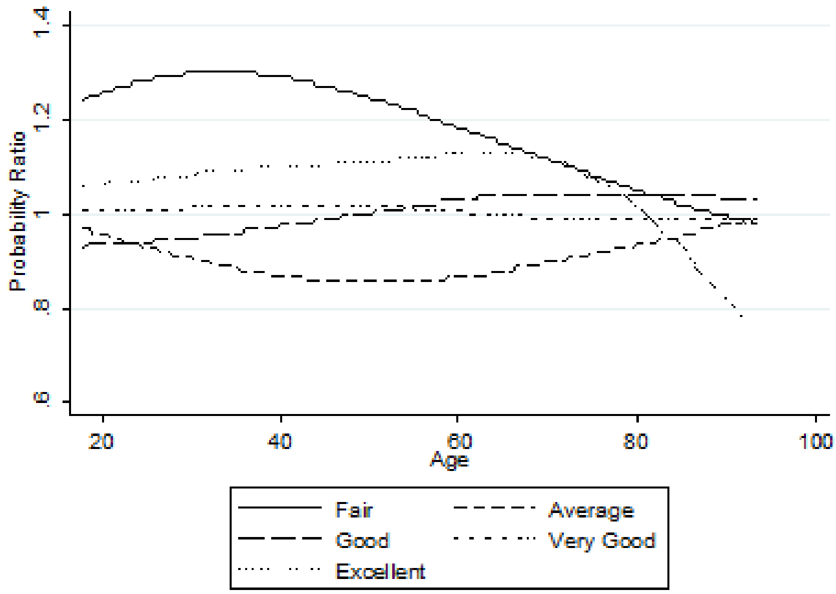

Figure 2: Ratio of Outcome Probabilities between Models IT5 and IT4

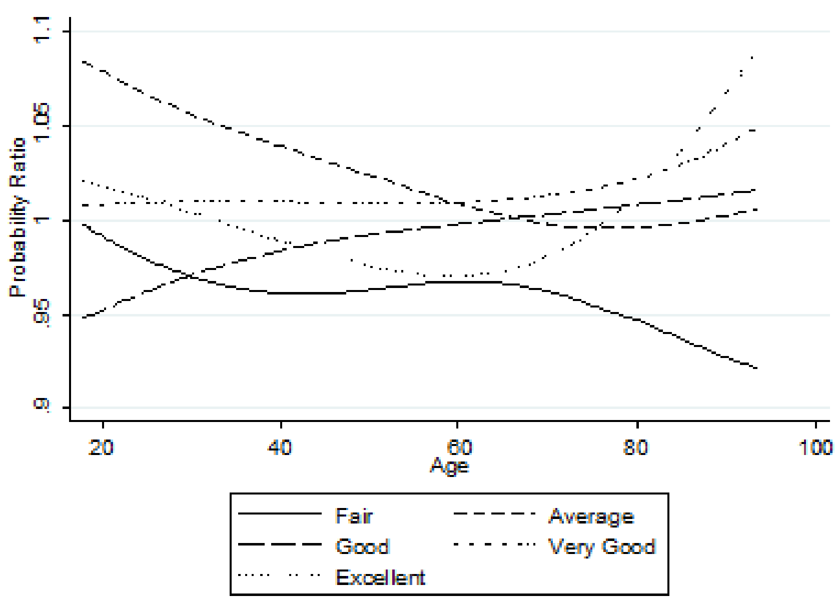

Figure 3: Ratio of Outcome Probabilities between Models IT6 and IT5

Allowing individual variances to be correlated to thresholds, it is seen that the selection of the most extreme categories of health is less likely between the ages of 30 and 70 years. *Fair* health is more likely to be selected by those under 70, and *very good* health is increasingly likely to be selected by older individuals. Each of these results has implications for economists whose key objective is to accurately measure the probability of certain responses.

6 Summary and Conclusions

This paper has introduced two new models that help to incorporate unobserved (or unobservable) heterogeneity into the thresholds and variance of ordered choice models. On the basis of a lack of appropriate threshold variables in many contexts, correlations were drawn between the threshold components of a parametrically-specified stochastic thresholds model, which is outlined in Greene and Hensher (2010) and further discussed in Weterings et al. (2011), amongst other applications. An extension allowing correlations between the threshold components and the (potentially unobservable) variance component was also proposed. Monte Carlo simulations were used to assess the performance of the models with correlated threshold components in small samples, as well as the performance of this model under misspecification.

Results indicate that allowing correlations between thresholds frequently results in additional flexibility, the power of the likelihood ratio statistic is questionable. When the degree of integration is increased, correct rejection of the false null hypothesis improves, albeit to a sub-optimal level. Much larger improvements are made when sample sizes are increased, suggesting the use of these models might largely be limited to datasets where more information is available to better integrate out the unobservable characteristics. Consequently, in small samples the use of information criteria was suggested as a measure of model fit, as opposed for formal tests of significance; although if significance of an LR test was found, true correlations would by quite likely. In addition, when large enough samples are utilised, we refer to the asymptotic properties of maximum likelihood estimation to justify the use of likelihood ratio statistics. Under misspecification, correlations between thresholds were found to be almost essential, with the additional flexibility fitting the data generating processes better, and allowing more accurate estimation of the latent regression coefficients than the uncorrelated thresholds model.

Each of the models was then estimated using the HILDA dataset, with an application to self-assessed health. Results indicated that both the correlated thresholds and correlated variance extensions were justified via the inspection of likelihood ratios. Importantly, there were differences in the conclusions drawn about latent regression covariates from each of the models. Differences were also found with regards to parameters related to unobserved components. For example, while individual variances were not significant in conjunction with individual thresholds, correlations between the variance and threshold components were found to be significant, resulting in a good improvement in the log-likelihood between the two models.

Post-estimation analysis investigated other model outputs that may have been affected by unobserved threshold/variance specifications. Marginal effects were seen to show significant change across two dimensions - parameter specific, and outcome specific. Probability plots also showed changes across models, with differences in estimated probabilities being substantial in some cases, and suggesting that, if prediction of model outcomes is important, allowing for such heterogeneity will be important in practice.

Overall, as a method for accounting for unobserved factors in the evaluation of outcome probabilities, correlated threshold and variance models provide substantial benefits in flexibility over their more restrictive uncorrelated counterparts. Each of the models has the advantage of only requiring additional computing power, and not additional variables, so can be applied when suitable threshold covariates or vignette questions are not available. If suitable covariates do exist, then use of these covariates in deterministic contexts might be more appropriate than use of these models. Alternatively, one could use such variables in addition to individual components in order to provide better model fit, and better control for reporting heterogeneity. Such cases are left for future research.

References

Y Benyamini, E Leventhal, and H Leventhal. Self-assessment of health: what do people know that predicts their mortality? *Research on Aging*, 21:477–500, 1999.

S Boes and R Winklemann. Ordered response models. *Allgemeines Statistisches Archiv*, 90(2):165–180, 2006.

S Cameron and J Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106:262–333, 1998.

F Cunha and S Navarro. The identification and economic content of ordered choice models with stochastic thresholds. University College Dublin, Gery Institute, Discussion Paper WP/26/2007, 2006.

P Garcia-Gomez, A Jones, and N Rice. Health effects on labour market exits and entries. *Labour Economics*, 17:62–76, 2010.

W Greene. *Limdep Version 9.0: Reference Guide*. Plainview, New York, Econometric Software Inc., 2007.

W Greene and D Hensher. *Modeling Ordered Choices*. Cambridge University Press, 2010.

C Hernndez-Quevedo, A Jones, and N Rice. Reporting bias and heterogeneity in self-assessed health. evidence from the british household panel survey. *Cuadernos Econmicos de ICE*, 75:63–97, 2008.

E Idler and Y Benyamini. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour*, 38(1):21–37, 1997.

G King, C Murray, J Salomon, and A Tandon. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98:191–207, 2004.

A Lecluyse and I Cleemput. Making health continuous: implications of different methods on the measurement of inequality. *Health Economics*, 15:99–104, 2006.

J Lemp, K Kockelman, and A Unnikrishnan. Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis and Prevention*, 43:370–380, 2011.

J Litchfield, B Reilly, and M Veneziani. An analysis of life satisfaction in albania: A heteroskedastic ordered probit model approach. Accepted for publication, 2010.

I Losonscz. Personality traits in hilda. HILDA survey research conference 2007, University of Melbourne, 2004.

C Pfeifer and T Cornelissen. The impact of participation in sports on educational attainment - new evidence from germany. *Economics of Education Review*, 29:94–103, 2010.

Powdthavee. Ill-health as a household norm: Evidence from other peoples health problems. *Social Science and Medicine*, 68:251–259, 2009.

M Pudney and M Shields. Gender, race, pay and promotion in the british nursing profession: Estimation of a generalized ordered probit model. *Journal of Applied Econometrics*, 15:367–399, 2000.

N Ritter and C Vance. The determinants of bicycle helmet use: Evidence from germany. *Accident Analysis and Prevention*, 43:95–100, 2011.

J Terza. Ordered probit: A generalization. *Communications in Statistics A. Theory and Methods*, 14:1–11, 1985.

T Weterings, W Greene, M Harris, and B Hollingsworth. Heterogeneity in ordered choice models: A review with applications to self-reported health. Unpublished paper available online at ..., 2011.

N Ziebarth. Measurement of health, health inequality, and reporting heterogeneity. *Social Science and Medicine*, 71:116–124, 2011.

Appendix 1. Variable descriptions

Table 10: Descriptive statistics and definitions for model covariates

Variable	**Average**	**St. Dev**	**Min**	**Max**	**Description**
SAH	3.361	0.954	1.000	5.000	Self-Assessed Health Variable
Married	0.629	0.483	0.000	1.000	1 if Married, 0 otherwise
Separated	0.139	0.346	0.000	1.000	1 if Separated, 0 otherwise
Immigrant	0.209	0.407	0.000	1.000	1 if Immigrant, 0 otherwise
Postgraduate	0.032	0.176	0.000	1.000	1 if highest level of education is Postgraduate, 0 otherwise
Bachelors	0.127	0.333	0.000	1.000	1 if highest level of education is a Bachelors degree , 0 otherwise
Other Tertiary	0.346	0.476	0.000	1.000	1 if highest level of education is another tertiary qualification, 0 otherwise
Year 12	0.144	0.351	0.000	1.000	1 if highest level of education is Year 12, 0 otherwise
Household Size	2.882	1.468	1.000	13.000	Number of individuals living in the household
Children 0-4	0.136	0.342	0.000	1.000	1 if have children between 0 and 4 years, 0 otherwise
Children 5-9	0.145	0.352	0.000	1.000	1 if have children between 5 and 9 years, 0 otherwise
Children 10-14	0.188	0.391	0.000	1.000	1 if have children between 10 and 14 years, 0 otherwise
Ln(Income)	0.780	0.656	0.000	8.176	Ln(Total Household Income+1)/100000
Age	0.440	0.180	0.150	0.930	Age of respondent (in 100 years)

Appendix 2. Self-Assessed Health across HILDA waves

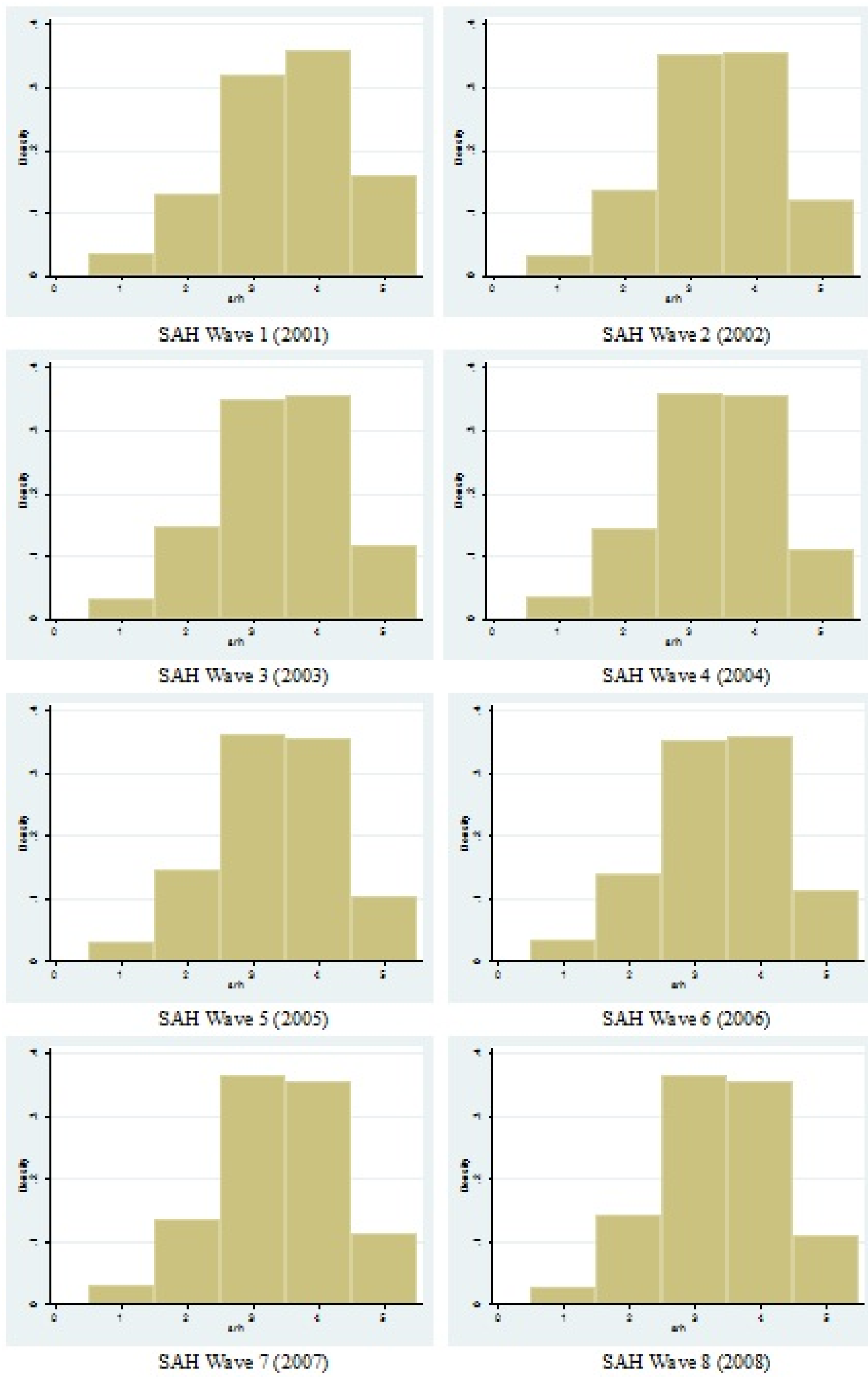

Figure 4: Self-Assessed Health Across HILDA waves

Appendix 3. Monte Carlo Simulation Results

Table 11: Set 1 - N = 500, M = 100, T = 8

Measure	**Model**	$\sigma_{\mathbf{Z_1Z_2}} = \mathbf{0}$	**0.3**	**0.7**	**-0.7**
β_1 **MSE**	OP	2.1E-05	1.4E-05	2.4E-06	1.4E-05
	UIT	2.3E-06	2.8E-06	4.3E-06	3.3E-06
	CIT	4.9E-06	4.6E-06	4.9E-06	4.3E-06
β_2 **MSE**	OP	2.4E-04	1.6E-04	2.3E-05	1.7E-04
	UIT	2.4E-05	3.2E-05	5.1E-05	3.7E-05
	CIT	5.5E-05	5.4E-05	5.9E-05	5.1E-05
Log-Likelihood	True Model	-2488.4	-2514.4	-2607.0	-2537.5
	OP	-2794.7	-2808.6	-2799.8	-2724.3
	UIT	-2677.1	-2696.7	-2741.8	-2677.0
	CIT	-2676.5	-2695.9	-2740.6	-2676.2
LR Test (% Rejected)	UIT vs OP	100%	100%	100%	100%
	CIT vs UIT	8%	10%	21%	13%

Table 12: Set 2 - N = 500, M = 200, T = 8

Measure	**Model**	$\sigma_{\mathbf{Z_1Z_2}} = \mathbf{0}$	**0.3**	**0.7**	**0.3, M=500**
β_1 **MSE**	OP	2.3E-05	1.5E-05	2.6E-06	1.3E-05
	UIT	1.5E-06	2.3E-06	4.5E-06	1.8E-06
	CIT	4.2E-06	4.4E-06	4.4E-06	4.1E-06
β_2 **MSE**	OP	2.7E-04	1.7E-04	2.5E-05	1.6E-04
	UIT	1.5E-05	2.6E-05	5.2E-05	1.8E-05
	CIT	4.8E-05	5.2E-05	5.1E-05	4.6E-05
Log-Likelihood	True Model	-2489.9	-2515.7	-2606.5	-2517.2
	OP	-2795.7	-2808.6	-2800.1	-2811.6
	UIT	-2677.9	-2696.5	-2741.2	-2698.4
	CIT	-2677.3	-2695.8	-2740.2	-2697.8
LR Test (% Rejected)	UIT vs OP	100%	100%	100%	100%
	CIT vs UIT	7%	11%	17%	7%

Table 13: Set 4 - N = 500, M = 100, T = 4

Measure	**Model**	$\sigma_{\mathbf{Z_1 Z_2}} = \mathbf{0}$	**0.3**	**0.7**
β_1 **MSE**	OP	2.2E-05	1.5E-05	3.5E-06
	UIT	2.7E-06	3.4E-06	5.0E-06
	CIT	8.7E-06	9.3E-06	9.6E-06
β_1 **MSE**	OP	2.4E-04	1.5E-04	2.4E-05
	UIT	2.0E-05	2.9E-05	4.8E-05
	CIT	9.5E-05	1.0E-04	1.0E-04
Log-Likelihood	True Model	-1414.4	-1429.2	-1480.6
	OP	-1583.1	-1589.1	-1584.7
	UIT	-1545.6	-1553.6	-1567.6
	CIT	-1545.0	-1552.9	-1566.8
LR Test (% Rejected)	UIT vs OP	100%	100%	100%
	CIT vs UIT	8%	11%	13%