Bayesian Estimation of A Random Effects Heteroscedastic Probit Model

Yuanyuan Gu∗
School of Economics
University of New South Wales

Denzil G Fiebig
School of Economics
University of New South Wales

Edward Cripps
Department of Probability & Statistics
University of Sheffield

Robert Kohn
School of Economics
University of New South Wales

This Version: June 26, 2007

Abstract

A Bayesian analysis is given of a random effects binary probit model that allows for heteroscedasticity. Real examples and simulations illustrate the approach and show that ignoring heteroscedasticity when it is there leads to biased estimates and poor prediction. The computation carried out by Markov chain Monte Carlo simulation is made efficient by generating the parameters in blocks. We use the Deviance Information Criterion for model comparison and evaluate the predictive ability of different models using the Receiving Operating Characteristic curve.

Keywords: Deviance Information Criterion; Receiving Operating Characteristic curve
JEL Classification: C11, C23, C25

∗Corresponding author. E-mail: yuanyuan.gu@student.unsw.edu.au
1 Introduction

It is well known in the literature that fitting a homoscedastic probit regression model when the data generating process is heteroscedastic may result in biased and inconsistent parameter estimates and predictions; see, for example, Parikh and Sen (2006). This finding also extends to more complex probit models such as the ordered probit model and the bivariate probit model; see, for example, Ritsila and Haapanen (2003) and Alpu and Kurt (2004). Routines for these models have already become standard in statistical software such as Stata and Limdep.

Many efforts have also been made to deal with heteroscedasticity in the context of linear regression with random effects. Li and Stengos (1994) derived an adaptive estimator when heteroscedasticity was placed on the remainder error term, i.e. the error term for the individuals within groups. Roy (2002) provided a similar adaptive estimator to model heteroscedasticity on the group specific error, i.e. the error term representing the random effects. Most recently Baltagi et al. (2005) compared these two models in terms of the sensitivity to misspecification of the form of heteroscedasticity. See also Baltagi (2005) for a review of this literature.

Although there is extensive literature on the heteroscedastic probit model and the heteroscedastic regression with random effects, to the best of our knowledge, so far no research has been done to tackle their combination. The purpose of this paper is to construct a random effects heteroscedastic probit (REHET) model. The random effects probit (REP) model is simply an extension to the standard probit model but takes into account the unobserved, group specific heterogeneity by adding random intercepts. Similarly, the homoscedasticity assumption needs to be tested if the REP model is used. Unfortunately, it has always been ignored by applied economists probably due to the difficulty in estimating the REHET model. For example, Anantham (2006) realized the existence of heteroscedasticity in his panel binary data when he was analysing strategic behavior in Blackjack. However, because of the difficulty in estimating such a complicated model, he chose to ignore the random effects and simply estimated a heteroscedastic probit model as a treatment. This paper motivates our REHET model which provides practitioners with an attractive framework for estimation and inference.
Unlike linear regression models, there is no standard estimation methodology for non-linear models. The REP model usually assumes that the random effects are normally distributed, thus numerical integration methods such as quadrature (Butler and Moffitt (1982)) or stochastic integration method maximum simulated likelihood introduced by Keane (1993) have been used to estimate the model.

Similarly to Li and Stengos (1994), our article allows heteroscedasticity on the remainder error term variance. Quadrature procedure and the method of maximum simulated likelihood might work on this model but here a fully Bayesian analysis is implemented. Instead of integrating over the normal distribution of the random intercepts, Bayesian analysis only needs to calculate posterior marginals. We will show how to do so using Markov chain Monte Carlo techniques (see for example, Chib and Greenberg (1995)). Furthermore, we can calculate marginal effects and carry out Bayesian inference. The REP model and REHET model are then compared using the Deviance Information Criterion (DIC) introduced by Spiegelhalter et al. (2002) while their predictive ability are evaluated by plotting the Receiving Operating Characteristic curve (ROC curve). Details are given in the appendix that describe how to ensure the sampling scheme is efficient.

2 The Model

2.1 Random Effects Probit (REP)

We start with the latent variable specification of the standard random effects probit model in which the random effect results from a group specific error term, i.e.

\[ y_{it}^* = x_{it} \beta + \mu_i + \nu_{it}, \]  

for groups \(i=1,...,N\) and individuals \(t=1,...,T_i\) in each group. \(x_{it}\) is a \(1 \times k\) vector of explanatory variables and \(\beta\) is a \(k \times 1\) vector of regression coefficients. The random effects \(\mu_i\) are assumed to be independent of the covariates \(x_{it}\) and the remainder error term \(\nu_{it}\). The sign of the latent variable determines the observed binary outcome variable, i.e.
We assume that both $\mu_i$ and $\nu_{it}$ are independent, identically distributed, and normal, i.e.

\[
\begin{align*}
\mu_i &\sim N(0, \sigma^2_{\mu}), \\
\nu_{it} &\sim N(0, \sigma^2_{\nu}).
\end{align*}
\]

For the standard random effects probit, we set $\sigma^2_{\nu} = 1$ without loss of generality.

## 2.2 Random Effects Heteroscedastic Probit (REHET)

We now introduce a model to include heteroscedasticity in the remainder error term $\nu_{it}$. The idea is adapted from models for heteroscedasticity in linear regression as described by Yau and Kohn (2003) and Chan et al. (2006):

\[
\begin{align*}
\nu_{it} &\sim N(0, \sigma^2_{\nu}), \\
\sigma^2_{\nu} &= \exp(\delta_0)\exp(w_{it}\delta) = \sigma^2\exp(w_{it}\delta),
\end{align*}
\]

where $w_{it}$ is a vector of covariates which hopefully explain the heterogeneity and $\delta$ is a $m \times 1$ vector of coefficients. Often the information available to explain the variance function is the same as, or, a subset of, those in $x_{it}$. To ensure the model is identified set $\sigma^2 = 1$ or $\delta_0 = 0$ and make sure $w_{it}$ does not have an intercept.

For the convenience of the Bayesian analysis in the next section, we write the REHET model as

\[
Y^* = Z\alpha + \nu,
\]

where
• $Y^* = (y^*_1, \ldots, y^*_{i_1}, \ldots, y^*_{N_1}, \ldots, y^*_{N_T N})'$ is a $(\sum_{i=1}^N T_i) \times 1$ vector of latent variable responses.

• Let $1_{T_i}$ denote a $T_i \times 1$ vector of ones and let $X_i = (x'_{i_1}, \ldots, x'_{iT_i})'$. Then

$$Z = \begin{bmatrix} 1_{T_1} & 0 & 0 & X_1 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 1_{T_N} & X_N \end{bmatrix}$$

which is a $(\sum_{i=1}^N T_i) \times (N + k)$ matrix. For convenience, denote

$$Z = \begin{bmatrix} J & X \end{bmatrix}$$

where $X = (X'_1, \ldots, X'_N)'$ and $J$ is defined conformally.

• $\alpha = [U', \beta']'$ where $U = (\mu_1, \mu_2, \ldots, \mu_N)'$ so that $\alpha$ is a $(N + k) \times 1$ vector.

• $\nu$ is a $(\sum_{i=1}^N T_i) \times 1$ error vector with mean zero and covariance a $(\sum_{i=1}^N T_i) \times (\sum_{i=1}^N T_i)$ diagonal matrix. Write its diagonals as a $(\sum_{i=1}^N T_i) \times 1$ vector, i.e.

$$\Sigma = (\sigma^2_{11}, \ldots, \sigma^2_{i_1T_i}, \ldots, \sigma^2_{N_1}, \ldots, \sigma^2_{N_T N})'$$

and assume

$$\log(\Sigma) = W \delta \quad (8)$$

where $W = (W'_1, \ldots, W'_N)'$ and $W_i = (w'_{i_1}, \ldots, w'_{iT_i})'$.

3 Bayesian Analysis

3.1 Prior specification

Given the model specified in section 2.2, we now discuss the priors for the parameters. Following Chan et al. (2006), for a given value of $\delta$, let
\[ D(\delta) = \text{diag}(\exp(w_{11}\delta/2), \ldots, \exp(w_{1T_1}\delta/2), \ldots, \exp(w_{N1}\delta/2), \ldots, \exp(w_{NT_N}\delta/2)) \]

and define \( \tilde{Y}^* = D(\delta)^{-1}Y^* \) and \( \tilde{Z} = D(\delta)^{-1}Z \). Apparently,

\[
\tilde{Z} = \begin{bmatrix} \tilde{J} & \tilde{X} \end{bmatrix}
\]

where \( \tilde{J} = D(\delta)^{-1}J \) and \( \tilde{X} = D(\delta)^{-1}X \).

Then (7) can be written as,

\[
\tilde{Y}^* = \tilde{Z}\alpha + \zeta, \quad \zeta \sim N(0, I).
\]

For a given value of \( \delta \), the prior for \( \beta \) is

\[
\beta | \delta \sim N(0, c_{\beta}(\tilde{X}^T\tilde{X})^{-1}).
\]

The prior was called a g-prior by Zellner (1986); please also see Smith and Kohn (1996) who interpret it as a shifted and inflated version of the likelihood. The parameter \( c_{\beta} \) is a scale hyperparameter associated with a prior density and taken as a large positive integer such as the sample size. In our practice, we let it be the total number of observations, i.e. \( c_{\beta} = \sum_{i=1}^{N} T_i \).

For a given value of \( \sigma^2_\mu \), the prior for \( U \) is

\[
U | \sigma^2_\mu \sim N(0, \sigma^2_\mu I_N).
\]

The prior for \( \alpha \), given \( \sigma^2_\mu \) and \( \delta \), is

\[
\alpha | \sigma^2_\mu, \delta \sim N(0, V_\alpha),
\]

where
which is a \((N + k) \times (N + k)\) matrix.

We now consider the parameters in the variance function. The prior for \(\delta\) is

\[
\delta \sim N(0, c_\delta I_m)
\]

where \(c_\delta\) is the scale hyperparameter.

Finally, we choose the same inverse gamma prior \(\text{IG}(a, b)\) for both the scale parameters \(c_\delta\) and \(\sigma_\mu^2\), with \(a = 1 + 10^{-10}\) and \(b = 1 + 10^{-5}\), which is proper but highly noninformative.

### 3.2 Sampling scheme

This section describes a Markov chain Monte Carlo sampling scheme for generating \(\alpha, \delta, \sigma_\mu^2\) and \(c_\delta\). Technical details are given in Appendix A.

The sampling scheme consists of an initialization step followed by 5 steps that are repeated at each iteration. After a sufficiently large number of iterations, called the ‘warmup’ or ‘burnin’ period, the iterates are assumed to be generated from the joint posterior distribution and it is these iterates that are used for inference.

The sampling scheme proceeds as follows:

1. Initialize \(\alpha, \delta, \sigma_\mu^2\) and \(c_\delta\).
2. Apply Albert and Chib (1993) type approach to generate latent variables \(Y^*\) from truncated normal distributions.
3. Generate \(\alpha\) in blocks from its full conditional distribution which is a normal density.
4. Generate \(\delta^p\) from a student-t approximation to \(p(\delta^p|Y^*, \alpha, c_\delta, \sigma_\mu^2)\). The proposed value \(\delta^p\) is then accepted or rejected according to the Metropolis-Hastings rule.
5. Generate $\sigma^2_\mu$ from its full conditional distribution which is an inverse gamma distribution.

6. Generate $c_3$ from its full conditional distribution which is an inverse gamma distribution.

3.3 Posterior inference

In order to confidently use the post burnin iterates for inference, it is necessary to check that the sampling scheme has converged. We judge convergence visually by running the sampling scheme from three different initial positions and plot various functionals of the iterates on the same graph. Successful convergence is indicated by the overlap of the functionals from the three chains. In all our examples we used a burnin period of 10,000 iterations and we then used the next 10,000 iterates for inference.

In Bayesian inference, the parameters are usually estimated by their posterior means. A desirable outcome of any model is an understanding how small changes in the covariates affect the mean response. In a homoscedastic linear regression model, such marginal effects are given by the regression coefficients, but this is not the case in the heteroscedastic probit regression model. There has been a considerable discussion in the literature on how to compute such marginal effects; see for example Greene (2003) (p.668). In general, it is possible to evaluate the expressions (derivatives for continuous variables and differences between two probabilities for dummy variables) at the sample means of the data or evaluate the marginal effects at every observation and use the sample mean of the individual marginal effects. We will use the second method, i.e. averaging the individual marginal effects, which is favored in current practice.

Given all the parameters generated from $j$th iteration, the estimated individual marginal effect for a continuous variable $h$ that is in both the mean and the variance functions is
\[ \hat{m}_{e[i]}^{[j]} = \frac{\partial p(y_{it} = 1|\mu_i^{[j]}, \beta^{[j]}, \delta^{[j]}, x_{it}, w_{it})}{\partial h_{it}} \]
\[ = \phi \left( \frac{\mu_i^{[j]} + x_{it}\beta^{[j]}}{\exp(\frac{w_{it}\delta^{[j]}}{2})} \right) \frac{\beta_h^{[j]} - (\mu_i^{[j]} + x_{it}\beta^{[j]})\delta_h^{[j]}}{\exp(\frac{w_{it}\delta^{[j]}}{2})}. \quad (14) \]

Only the first term applies if \( h \) appears only in the mean function.

Hence, in the \( j \)th iteration, the marginal effect for \( h \) is the average of the total individual marginal effects:

\[ \hat{m}_{e}^{[j]} = \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{m}_{e_{it}}^{[j]} \]
\[ \quad (15) \]
where \( T \) is the total number of observations, that is, \( T = \sum_{i=1}^{N} T_i \).

Finally, we average all the \( \hat{m}_{e}^{[j]} \) \( (j = 1, ..., S) \) to get the Bayesian estimate of the marginal effect:

\[ \hat{m}_{e} = \frac{1}{S} \sum_{j=1}^{S} \hat{m}_{e}^{[j]} \]
\[ \quad (16) \]
which is an estimate of the expression

\[ E(me|X, W) = \iiint \left( \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial p(y_{it} = 1|\mu_i, \beta, \delta, x_{it}, w_{it})}{\partial h_{it}} \right) p(U, \beta, \delta|X, W) dU d\beta d\delta. \quad (17) \]

When \( h \) is a dummy variable, we simply calculate two probabilities where \( h \) is first set to 0 and then to 1, and compute \( \hat{m}_{e_{it}}^{[j]} \) as the difference in probabilities.

To end this section, we show how to calculate predicted probabilities using the iterates. We note that

\[ E(p_{it}|x_{it}, w_{it}) = \iiint p(y_{it} = 1|\mu_i, \beta, \delta, x_{it}, w_{it}) p(\mu_i, \beta, \delta|x_{it}, w_{it}) d\mu_i d\beta d\delta, \quad (18) \]
which is estimated by
\[
\hat{p}_t = \frac{1}{S} \sum_{j=1}^{S} p(y_{it} = 1 | \mu_i^{[j]}, \beta^{[j]}, \delta^{[j]}, x_{it}, w_{it}).
\] (19)

4 Examples

4.1 Papsmear data

The first example is from a stated preference study of Australian women who choose whether or not to have a Pap test; see Fiebig and Hall (2005). There are 79 women in the sample and each respondent is presented with 32 scenarios. Thus, in terms of the panel structure described in Section 3, \(N = 79\) and \(T_i = 32\), which means that there are a relatively large number of repeated observations for estimation. Our second example will have a structure more common with observational studies, where \(T_i\) is relatively small compared to \(N\).

In each scenario the hypothetical test is described by different combinations of 5 attributes or variables whose definitions are given in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>knowgp</td>
<td>=1 if the GP is known to the patient; =0 otherwise</td>
</tr>
<tr>
<td>sexgp</td>
<td>=1 if the GP is male; =0 if the GP is female</td>
</tr>
<tr>
<td>testdue</td>
<td>=1 if the patient is due or overdue for a Pap test; =0 otherwise</td>
</tr>
<tr>
<td>drrec</td>
<td>=1 if GP recommends that the patient has a Pap test; = 0 otherwise</td>
</tr>
<tr>
<td>papcost</td>
<td>cost of test in $\text{A}$</td>
</tr>
</tbody>
</table>

Table 2 presents the results of estimating a random effects probit model (REP) and a random effects probit model with heteroscedasticity (REHET). Note that for the REHET model, the covariates selected into the variance function are the same as those in the mean function because we do not have prior information on the source of heteroscedasticity.

Table 2 shows that the estimated coefficients of the mean function substantially differ in terms of magnitude between REHET and REP models but their signs remain the same.
Table 2: Estimates of Papsmear data

<table>
<thead>
<tr>
<th></th>
<th>REP</th>
<th>REHET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est (Std.Er)</td>
<td>ME (Std.Er)</td>
</tr>
<tr>
<td>constant</td>
<td>-0.2888 (0.1822)</td>
<td>–</td>
</tr>
<tr>
<td>knowgp</td>
<td>0.3039 (0.0649)</td>
<td>0.0661 (0.0139)</td>
</tr>
<tr>
<td>sexgp</td>
<td>-0.6608 (0.0672)</td>
<td>-0.1466 (0.0142)</td>
</tr>
<tr>
<td>testdue</td>
<td>1.1891 (0.0763)</td>
<td>0.2764 (0.0166)</td>
</tr>
<tr>
<td>drrec</td>
<td>0.4967 (0.0736)</td>
<td>0.1105 (0.0163)</td>
</tr>
<tr>
<td>papcost</td>
<td>-0.0095 (0.0029)</td>
<td>-0.0021 (0.0006)</td>
</tr>
<tr>
<td>knowgp</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>sexgp</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>testdue</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>drrec</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>papcost</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\sigma_\mu$</td>
<td>1.3403 (0.1359)</td>
<td>–</td>
</tr>
</tbody>
</table>

The magnitudes of the estimated marginal effects, measuring the changes on the predicted probability due to the change on each independent variable, are almost the same for both models. However, the interpretation is somewhat different as part of the estimated marginal effect in the case of the REHET results is due to the influence of two significant effects in the estimated variance function. There is significantly less variability in women’s screening choices when their GP is male and when the test is due. Both of these results are sensible. On average, women are less likely to test when their GP is male but for some women the gender of their GP does not have a big impact leading to increased variability in their choices. While women are more likely to test when the test is due there is considerable debate about what the recommended screening interval should be which is likely to translate into more variability in testing choices.

It is also the case that the estimated marginal effects, which are calculated as averages over the sample, mask differences in the distribution over individuals of the estimated effects.
Boxplots of the individual marginal effects provide an excellent visual evidence of this variability. Figure 1 presents the boxplots for all the 5 independent variables, where we can see that although the median marginal effects are almost identical for the two models, variability grows substantially over the individual marginal effects by using REHET. Interestingly, this is not only true for sexgp and testdue but also for the other three variables. The reason is that the significant heteroscedasticity will affect the marginal effects for all the independent variables (see equation 14).

Finally, it is worth noting that although REP reports significant covariates for the mean function, these estimates ought to be treated with caution since REHET suggests that heterogeneity does exist and ignoring this may result in biased and inconsistent estimates. In the next section, we will compare the REP and REHET models for the Papsmear data using several criteria.

4.2 Model comparison

4.2.1 DIC

We first use the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) to choose between REP and REHET for the Papsmear data.

The DIC is a hierarchical modeling generalization of the Akaike information criterion (AIC) that is based on the posterior distribution of the deviance statistic

\[ D(\theta) = -2\log[p(y|\theta)] + 2\log[f(y)], \]

where \( y \) is the data, \( \theta \) represents the unknown parameters of the model, \( p(y|\theta) \) is the likelihood function, and \( f(y) \) is some standardizing function of the data only.

In this approach, the quality of a fit is measured by the posterior expectation of the deviance, \( \bar{D} = E_\theta(D(\theta)) \); the larger it is, the worse the fit. The complexity is measured by the effective number of parameters, \( p_D = \bar{D} - D(\bar{\theta}) \), where \( \bar{\theta} \) is the posterior expectation of \( \theta \). \( p_D \) is often less than the total number of unknown parameters in hierarchical models.

The DIC is then defined as \( \text{DIC} = \bar{D} + p_D \). The idea is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized not only by the value
Figure 1: Boxplots of individual marginal effects for knowgp, sexgp, testdue, drrec and papcost: REP results in column 1, REHET results in column 2
of $\bar{D}$ but also by the effective number of parameters. Since $\bar{D}$ will decrease as the number of parameters in a model increases, the $p_D$ term compensates for this effect by favoring models with a smaller number of parameters.

The advantage of DIC over other criteria such as Bayes factor, for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. By storing the iterates of both $\theta$ and $D(\theta)$ during MCMC sampling, we can simply estimate $\bar{D}$ by the sample mean of the generated values of $D$ and $D(\bar{\theta})$ by plugging in the sample mean of the simulated values of $\theta$.

In calculating DIC, we adopt the so-called null standardization criterion

$$D(\theta) = -2\log[p(y|\theta)]$$

as ignoring the standardizing term $2\log[f(y)]$ does not affect the conclusion. Table 3 reports the values of the DIC and related quantities for the two models: REP and REHET.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\bar{D}$</th>
<th>$D(\bar{\theta})$</th>
<th>$p_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>REP</td>
<td>1970.2</td>
<td>1892.5</td>
<td>77.7</td>
<td>2047.9</td>
</tr>
<tr>
<td>REHET</td>
<td>1885.2</td>
<td>1802.6</td>
<td>82.7</td>
<td>1967.9</td>
</tr>
</tbody>
</table>

$\bar{D}$ for the REHET model is smaller than that for the REP model, which indicates the former one fits the Papsmear data better. The table also shows that although the effective number of parameters, $p_D$, is greater for the REHET model than the REP model, the DIC value for the REHET model is still smaller than the DIC value for the REP model. Therefore, for the papsmear data the REHET model is preferred over the REP model when performance is based on DIC and we conclude that heteroscedasticity should not be ignored.

4.2.2 ROC curve

Another natural criterion to compare the REP and REHET models is their predictive performance. Section 3.3 shows that the predictive probabilities $\hat{p}_{it}$ can be easily computed.
However, the allocating a predicted value of 1 or 0 to the binary variable $y_{it}$ involves specifying a probability cutpoint such that if $\hat{p}_{it}$ lies above this cutpoint a value of 1 is allocated, otherwise a value of 0 is allocated. A common practice is to set the cutpoint 0.5; see for example Greene (2003). This may not be a good choice as the number of correct predictions for $y = 1$ and $y = 0$ could be notably different even when there is only a small change on the cutpoint. Having noticed this, we then use the ROC curve to examine the prediction performance for binary choice models. The Receiver Operating Characteristic curve (ROC curve) is a widely used technique in medical research which plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cutpoints of a diagnostic test. In our context, the sensitivity is $p(\hat{y} = 1|y = 1)$ and specificity $p(\hat{y} = 0|y = 0)$, and thus 1-specificity $p(\hat{y} = 1|y = 0)$. Given different cutpoints from $(0, 0.1, ..., 0.9, 1)$, we have different pairs of sensitivity and 1-specificity values. If the sensitivity is plotted against the 1-specificity, we then have a ROC curve. The ROC curve shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity). For each value of 1-specificity, we prefer a larger sensitivity. Alternatively, for each value of sensitivity we prefer a smaller 1-specificity. Figure 2 displays the ROC curves generated from the REP and REHET models for the Papsmear data and suggests that over all values of 1-specificity/sensitivity the REHET model either performs better or equivalently compared to the REP model.

### 4.3 Simulation from the fitted model

We use the estimates for the REP model fitted to the Papsmear data (see Table 2) to carry out the first simulation study. Denote the estimates of regression coefficients and the standard deviation of the random effects used to simulate the data $\beta$ (a $6 \times 1$ vector), $\delta$ (a $5 \times 1$ vector) and $\sigma_\mu$. One hundred replications of data are generated by

$$
y^*_it = x_{it}\beta + \mu_i + \nu_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T_i
$$

where $x_{it}$ is a $(1 \times k)$ vector of covariates, $\mu_i$ is the random effect for group $i$ generated from $N(0, \sigma^2_\mu)$ and $\nu_{it}$ is a Normal random deviate with mean zero and standard deviation $\sigma_{it}$.
which is set as 1.

We construct the binary response $y_{it}$ by setting

$$ y_{it} = \begin{cases} 
1 & \text{if } y_{it}^2 > 0 \\
0 & \text{otherwise.} 
\end{cases} $$

The data generated are then used to estimate both the REP model and the REHET model. As before, in each replication the posterior means are saved as the estimates and Table 4 contains the average and standard deviation of these estimates obtained from 100 replications.

The results in Table 4 show that the REP model and the REHET model perform similarly because the coefficient estimates of the variance function in the REHET model shrink to nearly zeros. It suggests that if heteroscedasticity is not there, we will not lose anything by using the REHET model.
### Table 4: Simulation 1 result

<table>
<thead>
<tr>
<th></th>
<th>Truth</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>Std</td>
<td>Est</td>
<td>Std</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₀</td>
<td>-0.2888</td>
<td>0.1740</td>
<td>-0.3436</td>
<td>0.1912</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₁</td>
<td>0.3039</td>
<td>0.0719</td>
<td>0.3199</td>
<td>0.0967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₂</td>
<td>-0.6608</td>
<td>0.0672</td>
<td>-0.6775</td>
<td>0.0994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₃</td>
<td>1.1891</td>
<td>0.0755</td>
<td>1.2326</td>
<td>0.1755</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₄</td>
<td>0.4967</td>
<td>0.0757</td>
<td>0.5205</td>
<td>0.1209</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₅</td>
<td>-0.0095</td>
<td>0.0031</td>
<td>-0.0097</td>
<td>0.0035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ₁</td>
<td>-</td>
<td>-</td>
<td>0.0026</td>
<td>0.1606</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ₂</td>
<td>-</td>
<td>-</td>
<td>0.0324</td>
<td>0.1559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ₃</td>
<td>-</td>
<td>-</td>
<td>-0.0140</td>
<td>0.1671</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ₄</td>
<td>-</td>
<td>-</td>
<td>0.0192</td>
<td>0.1780</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ₅</td>
<td>-</td>
<td>-</td>
<td>0.0008</td>
<td>0.0085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σₘ</td>
<td>1.3673</td>
<td>0.1234</td>
<td>1.4223</td>
<td>0.2196</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the second simulation study, the data are generated using the estimates from the REHET model fitted to the Papsmear data (see Table 2). The data generating process is the same as the one in the first simulation study except that \( \sigma_{it} \) is generated by \( \sigma_{it} = \exp(w_{it}\delta/2) \) where \( w_{it} \) is made equivalent to \( x_{it} \) by excluding the latter’s constant component.

The results in Table 5 illustrate that if heteroscedasticity is present the REHET model produces very accurate estimates in both the mean and variance functions. On the other hand, the REP model gives estimates very far from the truth which is caused by the scaling effects as the model assumes the variance to be 1. As described in previous sections, the REP model may provide acceptable marginal effects estimates even at the presence of heteroscedasticity but it cannot compete with the REHET model in terms of goodness of fit and predictive ability.
Table 5: Simulation 2 result

<table>
<thead>
<tr>
<th></th>
<th>Truth</th>
<th></th>
<th></th>
<th>REP</th>
<th></th>
<th></th>
<th>REHET</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>Std</td>
<td>Est</td>
<td>Std</td>
<td>Est</td>
<td>Std</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.0758</td>
<td>-0.2908</td>
<td>0.1550</td>
<td>-0.0870</td>
<td>0.0961</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.1533</td>
<td>0.3372</td>
<td>0.0682</td>
<td>0.1714</td>
<td>0.0455</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.3594</td>
<td>-0.6714</td>
<td>0.0809</td>
<td>-0.3743</td>
<td>0.0786</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.5397</td>
<td>1.1729</td>
<td>0.0833</td>
<td>0.5747</td>
<td>0.1092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.1869</td>
<td>0.4374</td>
<td>0.0786</td>
<td>0.1954</td>
<td>0.0542</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.0042</td>
<td>-0.0088</td>
<td>0.0032</td>
<td>-0.0043</td>
<td>0.0015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.1855</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1790</td>
<td>0.1648</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>-1.1495</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-1.1256</td>
<td>0.2040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>-1.2914</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-1.2215</td>
<td>0.2384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>-0.3124</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-0.3205</td>
<td>0.2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0009</td>
<td>0.0084</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_\mu$</td>
<td>0.6499</td>
<td>1.3379</td>
<td>0.1138</td>
<td>0.6877</td>
<td>0.1211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4 Exercise data

This section reanalyses the data of Lume and Fiebig (2006), and in particular their analysis of determinants of physical activity in Australia. This data was obtained from the 1995 National Health Survey (NHS) conducted by the Australian Bureau of Statistics (ABS). After exclusions, the dataset contains information from 20,234 individuals living in 11,938 households. The number of people in each household ranges from 1 to 6. Lume and Fiebig (2006) used a random effects probit model to measure the heterogeneity from different households. We also fit a random effects probit model to accommodate this dataset and using our notation, $N = 11,938$ and $T_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, ..., N$.

The dependent variable $y_{it}$ represents the level of exercise for each individual. If he or she exercises at a high or moderate level, $y_{it} = 1$; otherwise, $y_{it} = 0$. We note that the original dataset actually categorizes exercise into four levels: high, moderate, low and sedentary, and an ordered probit with random effects can also be estimated. In our paper, only the binary
choice model is considered.

There are 52 covariates excluding the constant, which include gender, age, income, etc; see Lume and Fiebig (2006) for further details. In Lume and Fiebig (2006), the random effects are statistically significant which suggests a household effect exists. We now relax the assumption of homoscedasticity made by Lume and Fiebig (2006) and use the 52 variables in the variance function. However, most of the 52 variables are statistically insignificant in the variance function. Thus we select those significant variables only in the variance function and estimate the REHET model again.

The table in Appendix B displays the estimated results of the REP model and the REHET model with 4 variables in the variance function. The selection of the 4 variables is based on the results of estimating a REHET model with all variables in where these 4 variables have estimates that are statistically significant in the variance function.

The DIC values presented in Table 6 reveal that the REHET model is better than the REP model. Figure 3 confirms this for the ROC curves of the two models.

<table>
<thead>
<tr>
<th>Model</th>
<th>( \bar{D} )</th>
<th>( D(\hat{\theta}) )</th>
<th>( p_D )</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>REP</td>
<td>21222</td>
<td>18819</td>
<td>2404</td>
<td>23626</td>
</tr>
<tr>
<td>REHET</td>
<td>20697</td>
<td>17998</td>
<td>2699</td>
<td>23396</td>
</tr>
</tbody>
</table>

Although the estimated coefficients of the mean function are very different for the two models, by and large the estimated marginal effects are similar to each other. As before, the interpretation of the results should take into account the variables in the estimated variance function. There are significant differences in Australians’ exercise choices when they have dependent children, were born in a country other than Australia, New Zealand, UK, Southern Europe, Western Europe and Asia, consume more alcohol weekly and reside in South Australia. The reason for the South Australia is not clear, but the other results are quite sensible. For example, on average, people are less likely to exercise when they have dependent children but there is considerable variability in their choices which could relate to different age patterns of the dependent children. Overall having dependent kids
has a negative effect on the exercise, as seen from the marginal effect of -0.0677 which is a somewhat larger effect than that found in the REP results. The REHET results reveal the significance of both the random effects and the presence of heteroscedasticity.

5 Conclusion

Our article proposes a Bayesian analysis on the random effects heteroscedastic probit model. The paper illustrates using both real and simulated data that ignoring heteroscedasticity may lead to biased estimates and inaccurate predictions. Model comparison is carried out using DIC and ROC analysis. Our sampling scheme is made efficient by blocking the parameters.
Appendix A: Sampling Scheme

Generating $Y^*$

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $y_{it}^*$ is generated from its full conditional density: $N(\mu_i + x_{it} \beta, \sigma_{it}^2)$ with support $[0, \infty]$ if $y_{it} = 1$ and with support $[-\infty, 0]$ if $y_{it} = 0$ where $\sigma_{it}^2 = \exp\{w_{it} \delta\}$.

Generating $\alpha$

$U$ and $\beta$ are jointly generated through $\alpha$ from its full conditional density: $N((\tilde{Z}'\tilde{Z} + V_{\alpha}^{-1})^{-1}\tilde{Z}'\tilde{Y}^*, (\tilde{Z}'\tilde{Z} + V_{\alpha}^{-1})^{-1})$.

We note that when $N$ is large updating $\alpha$ will be computationally time consuming because of the required cholesky decomposition of $(\tilde{Z}'\tilde{Z} + V_{\alpha}^{-1}) = \Delta$. We avoid this problem by taking advantage of the structure of $\Delta$. Write,

$$\Delta = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}$$

where $A$ is an $N \times N$ diagonal matrix, $B$ is a $N \times k$ matrix and $C$ is a $k \times k$ matrix. To solve the cholesky decomposition we require $LL' = \Delta$ where $L$ is a lower triangular matrix. Write

$$L = \begin{bmatrix} L_{11} & O \\ L_{21} & L_{22} \end{bmatrix}$$

Then $L_{11}L_{11}' = A$, $L_{11}L_{21}' = B$ and $L_{21}L_{21}' + L_{22}L_{22}' = C$. It is easy and fast to solve for $L_{11}$ which is a diagonal matrix with the $i$th elements equal to the square root of the $i$th elements of $A$, for $i = 1, \ldots, N$. This comprises the bulk of the matrix $\Delta$ when $N$ is large. Solving next for $L_{21}$ and $L_{22}$ is then straightforward.

Generating $\delta$

We use the Metropolis-Hastings method to generate $\delta$ because the full conditional density of $\delta$ is not recognized as a standard distribution. Our approach is based on a one-step Fisher
Scoring approximation, an idea suggested by Gamerman (1997). In our article, we adapt the version of this idea used in Chan et al. (2006). It avoids the cumbersome process of iterating to the mode at each iteration but usually at the cost of a little higher correlation among iterates. That is, using the one-step Fisher Scoring method, we may need more iterations to achieve convergence. However, in our case just one step may get you most of the way to the mode and get a good proposal at smaller computational expense. We now explain the steps in the algorithm.

1. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, calculate $e_{it} = (y_{it} - \mu_i - x_{it}\beta)^2$.

2. Denote $\sigma_{it}^2$ and $\delta^c$ for the current values of $\sigma_{it}^2$ and $\delta$. Calculate

   \[ \eta_{it}^c = w_{it}^c \delta^c + \frac{e_{it} - \sigma_{it}^c}{\sigma_{it}^2} \]

   and write $\eta^c = (\eta_{i1}^c, \ldots, \eta_{iT}^c, \ldots, \eta_{N1}^c, \ldots, \eta_{NT}^c)'$.

3. Calculate a mean and variance for a proposal density which is multivariate t with 4 degrees of freedom. That is, $\hat{\delta}^p = (\frac{2}{c_3} I + W'W)^{-1} W' \eta^c$ and $\Delta^p = (\frac{1}{c_3} I + \frac{1}{2} W'W)^{-1}$.

The proposed value of $\delta$ is generated by a multivariate t-density with 4 degrees of freedom $T_4(\hat{\delta}^p, \Delta^p)$ and denote this proposal density as $Q(\delta^c \rightarrow \delta^p)$. Write $\hat{\delta}^c$ and $\Delta^c$ for the proposal mean and covariance matrix when taking a step in the reverse direction, and $Q(\delta^p \rightarrow \delta^c)$ for the corresponding multivariate t-density with 4 degrees of freedom $T_4(\hat{\delta}^c, \Delta^c)$. We accept the proposal $\delta^p$ with probability

\[ \min \left\{ 1, \frac{p(Y^*|\beta, U, \delta^p, \sigma_{it}^2, c_3)}{p(Y^*|\beta, U, \delta^c, \sigma_{it}^2, c_3)} \exp\left\{ -\frac{1}{2c_3} \delta^p' \Delta^p \delta^p \right\} Q(\delta^p \rightarrow \delta^c) \frac{Q(\delta^c \rightarrow \delta^p)}{Q(\delta^p \rightarrow \delta^c)} \right\}. \]

**Generating $\sigma_{\mu}^2$**

$\sigma_{\mu}^2$ is generated from its full conditional density which is the inverse gamma distribution:

\[ \text{IG}\left( \frac{N+2a}{2}, \frac{\sum_{i=1}^{N} \mu_i^2 + 2b}{2} \right). \]
Generating $c_\delta$

c_\delta$ is generated from its full conditional density which is the inverse gamma distribution: $\text{IG}(\frac{m}{2} + a, \frac{\delta' \delta}{2} + b)$. 


## Appendix B: Estimates of Exercise data

<table>
<thead>
<tr>
<th></th>
<th>REP Est (Std.Er)</th>
<th>REP ME (Std.Er)</th>
<th>REHET Est (Std.Er)</th>
<th>REHET ME (Std.Er)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-0.4059(0.1617)</td>
<td>–</td>
<td>-0.4601 (0.1838)</td>
<td>–</td>
</tr>
<tr>
<td>age</td>
<td>-0.0251(0.0057)</td>
<td>-0.0075(0.0017)</td>
<td>-0.0283(0.0064)</td>
<td>-0.0070(0.0016)</td>
</tr>
<tr>
<td>agsq</td>
<td>0.0002(0.0001)</td>
<td>0.00005(0.00002)</td>
<td>0.0002(0.0001)</td>
<td>0.00004(0.00002)</td>
</tr>
<tr>
<td>female</td>
<td>-0.9735(0.1211)</td>
<td>-0.2763(0.0301)</td>
<td>-1.0041(0.1264)</td>
<td>-0.2429(0.0286)</td>
</tr>
<tr>
<td>agefemale</td>
<td>0.0318(0.0064)</td>
<td>0.0094(0.0019)</td>
<td>0.0317(0.0067)</td>
<td>0.0079(0.0017)</td>
</tr>
<tr>
<td>agesqfemale</td>
<td>-0.0003(0.0001)</td>
<td>-0.0001(0.00002)</td>
<td>-0.0003(0.0001)</td>
<td>-0.0001(0.00002)</td>
</tr>
<tr>
<td>single</td>
<td>0.2217(0.0381)</td>
<td>0.0659(0.0117)</td>
<td>0.2215(0.0450)</td>
<td>0.0557(0.0114)</td>
</tr>
<tr>
<td>kids</td>
<td>-0.2037(0.0409)</td>
<td>-0.0594(0.0120)</td>
<td>-0.6374(0.1139)</td>
<td>-0.0677(0.0118)</td>
</tr>
<tr>
<td>kidsfemale</td>
<td>-0.0899(0.0530)</td>
<td>-0.0263(0.0155)</td>
<td>-0.1519(0.0781)</td>
<td>-0.0365(0.0191)</td>
</tr>
<tr>
<td>depkid</td>
<td>0.1161(0.0611)</td>
<td>0.0343(0.0186)</td>
<td>0.0681(0.0691)</td>
<td>0.0169(0.0173)</td>
</tr>
<tr>
<td>ndepkid</td>
<td>-0.1959(0.0476)</td>
<td>-0.0558(0.0129)</td>
<td>-0.2092(0.0525)</td>
<td>-0.0511(0.0123)</td>
</tr>
<tr>
<td>exeng</td>
<td>0.0770(0.0416)</td>
<td>0.0225(0.0120)</td>
<td>0.0828(0.0447)</td>
<td>0.0203(0.0109)</td>
</tr>
<tr>
<td>english</td>
<td>0.0916(0.0581)</td>
<td>0.0263(0.0167)</td>
<td>0.0820(0.0640)</td>
<td>0.0199(0.0155)</td>
</tr>
<tr>
<td>muborn</td>
<td>0.0430(0.0369)</td>
<td>0.0127(0.0111)</td>
<td>0.0533(0.0440)</td>
<td>0.0131(0.0109)</td>
</tr>
<tr>
<td>seborn</td>
<td>-0.1080(0.0670)</td>
<td>-0.0311(0.0194)</td>
<td>-0.0954(0.0792)</td>
<td>-0.0230(0.0192)</td>
</tr>
<tr>
<td>weborn</td>
<td>0.0272(0.0822)</td>
<td>0.0092(0.0244)</td>
<td>0.0171(0.0919)</td>
<td>0.0045(0.0231)</td>
</tr>
<tr>
<td>asborn</td>
<td>-0.2929(0.0613)</td>
<td>-0.0812(0.0161)</td>
<td>-0.3635(0.0741)</td>
<td>-0.0848(0.0162)</td>
</tr>
<tr>
<td>otborn</td>
<td>0.0080(0.0581)</td>
<td>0.0019(0.0176)</td>
<td>0.1990(0.0662)</td>
<td>0.0143(0.0171)</td>
</tr>
<tr>
<td>tert</td>
<td>0.2293(0.0382)</td>
<td>0.0696(0.0117)</td>
<td>0.2832(0.0466)</td>
<td>0.0727(0.0117)</td>
</tr>
<tr>
<td>diploma</td>
<td>0.1419(0.0423)</td>
<td>0.0424(0.0128)</td>
<td>0.1697(0.0498)</td>
<td>0.0436(0.0127)</td>
</tr>
<tr>
<td>trade</td>
<td>0.0219(0.0292)</td>
<td>0.0064(0.0088)</td>
<td>0.0514(0.0349)</td>
<td>0.0130(0.0087)</td>
</tr>
<tr>
<td>unemploy</td>
<td>0.0284(0.0706)</td>
<td>0.0095(0.0210)</td>
<td>0.0369(0.0826)</td>
<td>0.0090(0.0209)</td>
</tr>
<tr>
<td>nilf</td>
<td>0.0004(0.0520)</td>
<td>0.0012(0.0154)</td>
<td>0.0300(0.0635)</td>
<td>0.0074(0.0159)</td>
</tr>
<tr>
<td>hours</td>
<td>-0.0067(0.0015)</td>
<td>-0.0020(0.0004)</td>
<td>-0.0074(0.0017)</td>
<td>-0.0018(0.0004)</td>
</tr>
<tr>
<td>longhrs</td>
<td>-0.0267(0.0408)</td>
<td>-0.0078(0.0118)</td>
<td>-0.0176(0.0492)</td>
<td>-0.0043(0.0122)</td>
</tr>
<tr>
<td>alcohol</td>
<td>0.0003(0.0001)</td>
<td>0.0001(0.00002)</td>
<td>0.0003(0.0001)</td>
<td>0.0001(0.00003)</td>
</tr>
<tr>
<td>smoke</td>
<td>-0.1481(0.0306)</td>
<td>-0.0438(0.0090)</td>
<td>-0.2011(0.0378)</td>
<td>-0.0501(0.0090)</td>
</tr>
<tr>
<td>quit</td>
<td>0.2268(0.0339)</td>
<td>0.0682(0.0101)</td>
<td>0.2891(0.0424)</td>
<td>0.0728(0.0102)</td>
</tr>
</tbody>
</table>

*to be continued...*
<table>
<thead>
<tr>
<th></th>
<th>REP</th>
<th></th>
<th></th>
<th>REHET</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est (Std.Er)</td>
<td>ME (Std.Er)</td>
<td>Est (Std.Er)</td>
<td>ME (Std.Er)</td>
<td>Est (Std.Er)</td>
<td>ME (Std.Er)</td>
</tr>
<tr>
<td>hlthex</td>
<td>0.8543(0.0731)</td>
<td>0.2638(0.0221)</td>
<td>0.9849(0.0863)</td>
<td>0.2566(0.0205)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hlthvg</td>
<td>0.6553(0.0710)</td>
<td>0.1934(0.0206)</td>
<td>0.7512(0.0826)</td>
<td>0.1881(0.0191)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hlthg</td>
<td>0.4186(0.0701)</td>
<td>0.1228(0.0207)</td>
<td>0.4823(0.0796)</td>
<td>0.1200(0.0194)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hlthf</td>
<td>0.2950(0.0738)</td>
<td>0.0888(0.0234)</td>
<td>0.3482(0.0841)</td>
<td>0.0893(0.0217)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>undwght</td>
<td>-0.1341(0.0370)</td>
<td>-0.0386(0.0104)</td>
<td>-0.1431(0.0420)</td>
<td>-0.0354(0.0099)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ovrwght</td>
<td>-0.0932(0.0252)</td>
<td>-0.0270(0.0073)</td>
<td>-0.1177(0.0294)</td>
<td>-0.0293(0.0073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wgtmiss</td>
<td>-0.2561(0.0405)</td>
<td>-0.0725(0.0107)</td>
<td>-0.2770(0.0459)</td>
<td>-0.0669(0.0105)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>recond</td>
<td>0.1286(0.0495)</td>
<td>0.0391(0.0151)</td>
<td>0.1681(0.0585)</td>
<td>0.0427(0.0149)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>itcond</td>
<td>0.0621(0.0427)</td>
<td>0.0187(0.0128)</td>
<td>0.0945(0.0497)</td>
<td>0.0238(0.0127)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bthcond</td>
<td>0.1594(0.0385)</td>
<td>0.0466(0.0111)</td>
<td>0.2004(0.0463)</td>
<td>0.0489(0.0110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inc</td>
<td>0.0029(0.0011)</td>
<td>0.0009(0.0003)</td>
<td>0.0036(0.0013)</td>
<td>0.0009(0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rich</td>
<td>-0.1546(0.0943)</td>
<td>-0.0441(0.0265)</td>
<td>-0.1835(0.1212)</td>
<td>-0.0434(0.0293)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>incmiss</td>
<td>-0.0130(0.0461)</td>
<td>-0.0034(0.0136)</td>
<td>-0.0288(0.0549)</td>
<td>-0.0072(0.0136)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sinc</td>
<td>0.0008(0.0010)</td>
<td>0.0002(0.0003)</td>
<td>0.0007(0.0013)</td>
<td>0.0002(0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>srich</td>
<td>-0.0321(0.1063)</td>
<td>-0.0079(0.0311)</td>
<td>-0.0731(0.1372)</td>
<td>-0.0165(0.0338)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sincmiss</td>
<td>0.0540(0.0359)</td>
<td>0.0160(0.0159)</td>
<td>0.0688(0.0638)</td>
<td>0.0167(0.0164)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vic</td>
<td>-0.0567(0.0364)</td>
<td>-0.0166(0.0105)</td>
<td>-0.0515(0.0433)</td>
<td>-0.0126(0.0108)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qld</td>
<td>-0.0359(0.0486)</td>
<td>-0.0102(0.0138)</td>
<td>-0.0367(0.0577)</td>
<td>-0.0084(0.0140)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sa</td>
<td>-0.0684(0.0401)</td>
<td>-0.0205(0.0114)</td>
<td>0.0134(0.0553)</td>
<td>-0.0175(0.0123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wa</td>
<td>-0.0029(0.0454)</td>
<td>-0.0005(0.0134)</td>
<td>-0.0093(0.0554)</td>
<td>-0.0020(0.0138)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tas</td>
<td>-0.0578(0.0581)</td>
<td>-0.0169(0.0169)</td>
<td>-0.0564(0.0709)</td>
<td>-0.0137(0.0172)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nt</td>
<td>0.1061(0.0516)</td>
<td>0.0316(0.0155)</td>
<td>0.1276(0.0644)</td>
<td>0.0322(0.0164)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>act</td>
<td>0.1065(0.0486)</td>
<td>0.0321(0.0146)</td>
<td>0.1332(0.0568)</td>
<td>0.0339(0.0149)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rural</td>
<td>0.0548(0.0296)</td>
<td>0.0162(0.0088)</td>
<td>0.0518(0.0352)</td>
<td>0.0129(0.0088)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>otqld</td>
<td>0.1204(0.0659)</td>
<td>0.0360(0.0202)</td>
<td>0.1498(0.0803)</td>
<td>0.0377(0.0207)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kids</td>
<td>–</td>
<td>–</td>
<td>1.1763(0.1894)</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>otborn</td>
<td>–</td>
<td>–</td>
<td>-0.9850(0.3120)</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alcohol</td>
<td>–</td>
<td>–</td>
<td>0.0011(0.0005)</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sa</td>
<td>–</td>
<td>–</td>
<td>-0.4240(0.1584)</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σμ</td>
<td>0.5495(0.0267)</td>
<td>–</td>
<td>0.6803(0.0399)</td>
<td>–</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: kids = 1 if have dependent children; otborn = 1 if born in other countries; alcohol is estimated weekly alcohol consumption (ml); sa = 1 if reside in South Australia
References


