THE DANISH
NATIONAL CENTRE
FOR SOCIAL RESEARCH

Anders Holm
Mads Meier Jæger
Morten Pedersen

16:2008 WORKING PAPER

UNOBSERVED HETEROGENEITY IN THE BINARY LOGIT MODEL WITH CROSS-SECTIONAL DATA AND SHORT PANELS: A FINITE MIXTURE APPROACH

RESEARCH DEPARTMENT OF SOCIAL POLICY AND WELFARE SERVICES

UNOBSERVED HETEROGENEITY IN THE BINARY LOGIT MODEL WITH CROSS-SECTIONAL DATA AND SHORT PANELS

A Finite Mixture Approach

Anders Holm
Mads Meier Jæger
Morten Pedersen

Social Policy and Welfare Services
Working Paper 16:2008

Unobserved Heterogeneity in the Binary Logit Model with Cross-Sectional Data and Short Panels: A Finite Mixture Approach

Anders Holm*, Mads Meier Jæger**, and Morten Pedersen***

November 2008

This paper proposes a new approach to dealing with unobserved heterogeneity in applied research using the binary logit model with cross-sectional data and short panels. Unobserved heterogeneity is particularly important in non-linear regression models such as the binary logit model because, unlike in linear regression models, estimates of the effects of observed independent variables are biased *even* when omitted independent variables are uncorrelated with the observed independent variables. We propose an extension of the binary logit model based on a finite mixture approach in which we conceptualize the unobserved heterogeneity via latent classes. Simulation results show that our approach leads to considerably less bias in the estimated effects of the independent variables than the standard logit model. Furthermore, because identification of the unobserved heterogeneity is weak when the researcher has cross-sectional rather than panel data, we propose a simple approach that fixes latent class weights and improves identification and estimation. Finally, we illustrate the applicability of our new approach using Canadian survey data on public support for redistribution.

* Department of Sociology, University of Copenhagen, Øster Farimagsgade 5, Building 16, DK-1014 Copenhagen K, Phone +45 3532 3236, e-mail: ah@soc.ku.dk,
** The Danish National Centre for Social Research, Herluf Trolles Gade 11, DK-1052 Copenhagen K, Phone +45 33 48 08 77, e-mail: mads@sfi.dk,
*** Department of Sociology, University of Copenhagen.

Keywords: binary logit model, unobserved heterogeneity, latent classes, simulation.

Word count: 6,164.

1. Introduction

Many outcome variables in quantitative political and social science research represent binary events. Individuals may vote or not vote; local governments may decide to outsource or not to outsource services; and voters may agree or not agree with a specific political statement. Researchers interested in estimating the probability that the event under study occurs typically employ binary choice models such as the binary logit or probit regression models.

Despite their popularity this type of non-linear regression models poses a special set of challenges to researchers. One of the most important challenges is bias from *unobserved heterogeneity*. Estimates of the effect of independent variables on the binary outcome will be biased if the researcher does not observe all the relevant independent variables that affect the outcome (Wooldridge 2002). Bias from unobserved heterogeneity is particularly important in non-linear regression models because, unlike linear regression models, estimates of the effect of independent variables will be biased *even* if the unobserved heterogeneity is not correlated with the observed independent variables (Bretagnolle and Huber-Carol 1988; Abramson et al. 2000; Ejrnæs and Holm 2006).

Unobserved heterogeneity can be dealt with in a number of ways. If the researcher has panel data with repeated observations on the binary outcome of interest, unobserved heterogeneity is typically dealt with either by conditioning on the unobserved heterogeneity through random effects or by transforming the data to eliminate individual-specific fixed effects (see Halaby 2004). These methods reduce the potential parameter bias from unobserved heterogeneity. However, in many cases the researcher does not have panel data and relies on cross-sectional data with only one record for each observational unit. Alternatively, the researcher might have a short panel with only two records per observational unit. In these scenarios it is difficult to deal effectively with potential bias from unobserved heterogeneity because there is only little information in the data that allows the researcher to identify and correct for the unobserved heterogeneity.

This paper proposes a new approach to dealing with unobserved heterogeneity in the binary logit model with cross-sectional data and short panels. Our approach is designed to reduce bias from unobserved heterogeneity in applied research and builds on a finite mixture binary logit (FMBL) framework in which we approximate the unobserved heterogeneity component non-parametrically via latent classes. The latent classes capture latent sub groups in data that differ with regard to experiencing the binary outcome (see Goodman 1974; McCutcheon 1987). By conditioning on the latent classes in the FMBL model, we control for the unobserved heterogeneity and reduce bias in the parameter estimates of the effects of the observed independent variables. Evidence from most applied research suggests that only a small number of latent class is required to capture the unobserved heterogeneity (e.g., Heckman and Singer 1984; Davies 1993; Holm 2002).

The major challenge we face is to control adequately for unobserved heterogeneity when the data is not very informative about the unobserved heterogeneity. This is especially the case when we have only cross-sectional data or short panels. We argue that our FMBL approach is preferable to the standard binary logit model in terms of reducing bias in the effects of observed independent variables even when the unobserved heterogeneity is weakly identified. Consequently, even when identification is weak our approach is better than not addressing unobserved heterogeneity. Furthermore, we propose a simple method for improving identification which fixes the weight of one or more latent class in the FMBL model. This method makes the FMBL model easier to estimate, and we show that fixing a latent class weight has only a neglible impact on the other parameters in the model. Finally, we argue that instead of fixing the latent class weight at an arbitrary value one can use an automated grid search to find the optimal weight. This model, which we label the Finite Mixture Binary Logit model with Fixed Weights (FMBLfw), performs well in simulations and could be a feasible alternative to the standard binary logit model when the researcher has only cross-sectional data.

We run a series of simulations to evaluate the performance of our FMBL approach relative to the standard binary logit model. We find that estimates of the effects of observed independent variables are considerably less biased in the FMBL model than in the standard binary logit model. This turns out also to be the case when we fix a latent class weights at a pre-defined value. Our simulations suggest that in the FMBL model the latent classes capture some of the unobserved heterogeneity in the data and, in doing so they reduce bias in the effects of the observed independent variables. This result has implications for applied research since, by using a relatively simple method, it is possible to estimate an extended version of the binary logit model which is more robust to unobserved heterogeneity than the standard logit model. Furthermore, we illustrate the applicability of the FMBL model using data on public support for redistributive policies in Canada.

To maintain expositional clarity the paper focuses on the logit model with only a binary outcome variable. However, our approach easily generalizes to more complex situations, for example multinomial models (e.g., Jæger and Holm 2007) or duration models (e.g., Vermunt 1997; Bearse et al. 2007). Recent research suggests that identification problems may be less severe in these situations because multinomial and duration models have more information in the dependent variables. Consequently, it is likely that our suggested approach will be at least as useful in these cases.

The paper proceeds as follows. Section 2 describes the finite mixture binary logit (FMBL) framework, bias in the standard binary logit model and identification of the FMBL with cross-sectional and panel data. Section 3 reports results from a simulation study which, first, illustrates bias in the standard binary logit model, second, highlights identification problems in the FMBL model and, third, shows why fixing a particular parameter in the FMBL may improve identification. In section 4 we use Canadian panel data on public support for redistribution to illustrate the applicability of our approach. Section 5 concludes.

2. Statistical Framework

2.1 The Finite Mixture Binary Logit Model

This section presents the idea behind the finite mixture binary logit (FMBL) model and compares this model to the standard binary logit model. The section also explains how the FMBL captures unobserved heterogeneity.

The FMBL model can be seen as an extension of the standard binary logit model which also includes a latent class model that captures the effect of unobserved variables on the binary outcome variable. The outcome variable is Y and takes the values $y = 0$ and $y = 1$. We formulate the FMBL model with J ($j = 1,\ldots,J$) latent classes as

$$P(Y=1 \mid \mathbf{x}) = \sum_{j=1}^{j=J} P(Y=1 \mid \mathbf{x}, \Xi = \varepsilon_j) P(\Xi = \varepsilon_j) = \sum_{j=1}^{j=J} \frac{\exp(\alpha + \boldsymbol{\beta}\mathbf{x} + \varepsilon_j) P(\Xi = \varepsilon_j)}{1 + \exp(\alpha + \boldsymbol{\beta}\mathbf{x} + \varepsilon_j)}, \quad (1)$$

where α is a constant term, $\mathbf{x}$ is a vector of independent variables, $\boldsymbol{\beta}$ is a corresponding row vector of regression coefficients, ε_j is the effect of the *j*'th latent class on the probability of observing *Y* = *1*, and $P(\Xi = \varepsilon_j)$ is the proportion of the population that belongs to the *j*'th latent class. The model parameters to be estimated are α, $\boldsymbol{\beta}$, ε_j, and $P(\Xi = \varepsilon_j)$. The FMBL model takes into account unobserved heterogeneity arising from omitted independent variables into account through the inclusion of latent classes. Conceptually, the unobserved heterogeneity can be thought of either as a true discrete distribution of unobserved heterogeneity or as an approximation to any unknown distribution of unobserved heterogeneity, discrete or continuous (Lindsay 1983a, 1983b). The latent class proportions $P(\Xi = \varepsilon_j)$ must meet the restrictions: $P(\Xi = \varepsilon_j) > 0$ and $\sum_{j=1}^{j=j} P(\Xi = \varepsilon_j) = 1$. Hence, it is useful to re-parameterize the model when estimating the proportions $P(\Xi = \varepsilon_j)$ to

$$P(\Xi = \varepsilon_j) = \frac{\exp\left(\tilde{\delta}_j\right)}{\sum_{j=1}^{j=J} \exp\left(\tilde{\delta}_j\right)}, \qquad (2)$$

where, now, $\tilde{\delta}_j$, *j = 1,…J* are parameters to be estimated. Furthermore, we divide by $\tilde{\delta}_1$ to get

$$P(\Xi = \varepsilon_j) = \frac{\exp\left(\tilde{\delta}_j - \tilde{\delta}_1\right)}{1 + \sum_{j=2}^{j=J} \exp\left(\tilde{\delta}_j - \tilde{\delta}_1\right)} = \frac{\exp\left(\delta_j\right)}{1 + \sum_{j=2}^{j=J} \exp\left(\delta_j\right)} \qquad (3).$$

It follows from Equation (3) that the number of identifiable parameters for the latent class proportions is *J –1*. Furthermore, it also follows that re-defining $\tilde{\varepsilon}_j = \alpha + \varepsilon_j$ leaves $P(Y = 1 \mid \mathbf{x}, \Xi = \tilde{\varepsilon}_j) = P(Y = 1 \mid \mathbf{x}, \Xi = \varepsilon_j)$, *j = 1,…, J* and that we need to normalize one of the effects of the latent classes, ε_j. We use conventional dummy-coding and normalize $\varepsilon_1 = 0$.

In the following sections we present a simple version of the FMBL model with only one independent (continuous) variable and two latent classes. We use this simplified version to illustrate the intuition behind the FMBL model. It is conceptually straightforward to extend the model to situations with more independent variables and latent classes. In this simple version the model is written as

$$P(Y=1\mid x)=\sum_{j=1}^{j=2}\frac{\exp(\alpha+\beta x+\varepsilon_j)P(\Xi=\varepsilon_j)}{1+\exp(\alpha+\beta x+\varepsilon_j)}, \tag{4}$$

where x is a continuous independent variable and β is a regression coefficient, and where $\varepsilon_1=0$ and $\varepsilon_2=\varepsilon$. From Equation (4) we construct the log-likelihood function for a sample of n independent observations as

$$\ln L=\sum_{i=1}^{i=n} y_i \ln P(Y=1\mid x_i)+(1-y_i)\ln\left(1-P(Y=1\mid x_i)\right) \tag{5}$$

where

$$P(Y=1\mid x_i)=pP_{0i}+(1-p)P_{i\varepsilon}, \tag{6}$$

and

$$P_{0i}=\frac{\exp(\alpha+\beta x_i)}{1+\exp(\alpha+\beta x_i)}, \tag{7.1}$$

$$P_{i\varepsilon}=\frac{\exp(\alpha+\beta x_i+\varepsilon)}{1+\exp(\alpha+\beta x_i+\varepsilon)}, \tag{7.2}$$

and, finally, where $P(\Xi=0)=p$ and $P(\Xi=\varepsilon)=1-p$. Following Equations (2) and (3), in the case of only two latent classes the parameterization of the latent class probabilities is

$$P(\Xi=\varepsilon_1=0)=\frac{1}{1+\exp(\delta_2)}, \tag{8.1}$$

$$P(\Xi=\varepsilon_2)=\frac{\exp(\delta_2)}{1+\exp(\delta_2)}. \tag{8.2}$$

TABLE 1 HERE

TABLE 2 HERE

In the following example we illustrate how the FMBL and the standard binary logit model might lead to very different estimates of the effect of the observed independent variable x on the probability that $y = 1$. Consider Table 1 which uses stylized data. From the table we find that the log-odds ratio that $y = 1$ as opposed to $y = 0$ as a function of the independent variable x with four values is approximately 1. However, suppose that the frequency distribution in Table 1 is actually comprised from two latent sub groups with very different frequency distributions. The distribution of each sub group is shown in Table 2. Here, it is evident that in both sub groups the log-odds ratio that $y = 1$ as opposed to $y = 0$ is actually approximately 2. Consequently, if we ignore the latent sub groups in the data and estimate a standard binary logit model on the data in Table 1, we obtain an estimate of the log-odds ratio, β, of approximately 1. Since the actual log-odds ratio in each sub group is approximately 2, the bias in the estimated β is around 100 percent.

TABLE 3 HERE

We use the stylized data from Table 1 and estimate two regression models: A standard binary logit model and the FMBL with two latent classes to capture the two latent sub groups in the data. Results from these models are shown in Table 3. The estimate of β in the binary logit model is 1.175 which fits the frequency distribution in Table 1. By contrast, the estimate of β in the FMBL with two latent classes is 2.026 and replicates the frequency distributions in Table 2. Assuming that the frequency distributions were generated according to Table 2, the binary logit model yields very biased estimates of β. Interestingly, even though the two models give very different estimates of β model fit according to the log-likelihood is very similar, and the ratio of the log-likelihoods of the two models is only 1.003.

FIGURE 1 HERE

To illustrate the similarities between the binary logit and the FMBL models Figure 1 plots the predicted probabilities of $Y = 1$ obtained from the two models. From the figure it is clear that, despite very different estimates of β , there are only marginal differences between the predicted probabilities of the binary logit model and the FMBL model (which in this case yields perfect fit to the data because it represents a saturated model). It is likely that the variation in *x* will only yield minor discrepancies in predicted probabilities between the binary logit and the FMBL model. Furthermore, it will often be difficult to determine whether these discrepancies are due to non-linear effects of *x* on the log-odds of *Y* or due to the presence of unobserved heterogeneity, as captured via the latent classes.

This result shows that the regression coefficient may be severely biased when unobserved heterogeneity is present *even* when the heterogeneity is uncorrelated with the observed independent variables. In the following paragraphs we show why this is the case in the standard binary logit model but not in the linear regression model.

Consider a linear regression model in which the constant term depends on which of two latent classes an individual belongs to. For a fixed *x* we get

$$y_1 = \alpha_1 + \beta x + e, \qquad (9.1)$$

$$y_2 = \alpha_2 + \beta x + e, \qquad (9.2)$$

where *e* is the idiosyncratic error term and where the two classes are distributed in the population with probability *p* and *1-p*. If class membership is unobserved, the observed *y* will be the average of the two *y*'s from each of the latent classes with respect to the distribution of the two classes

$$p(\alpha_1 + \beta x + e) + (1-p)(\alpha_1 + \beta x + e) = \tilde{\alpha} + \tilde{\beta} x + e \qquad (10)$$

with $\tilde{\alpha} = p\alpha_1 + (1-p)\alpha_1$ and $\tilde{\beta} = \beta$. Equation (10) is then another linear regression model with different intercept but with the same regression coefficient for *x* as in the two latent class regressions. Consequently, when the unobserved heterogeneity is not correlated with *x*, estimates of β in the population are unbiased estimates of the slope parameters in the two (latent) classes.

This result does not carry over to the binary logit model. Similar to the previous case, consider a binary logit model in which the constant term differs by latent class

$$P_1(y=1)=\frac{\exp(\alpha_1+\beta x)}{1+\exp(\alpha_1+\beta x)}, \qquad (11.1)$$

$$P_2(y=1)=\frac{\exp(\alpha_2+\beta x)}{1+\exp(\alpha_2+\beta x)}. \qquad (11.2)$$

If class membership is unobserved, the observed probability[1] that *y=1* will be the average of the two *y*'s from each of the latent classes, with respect to the distribution of the two classes

$$p\frac{\exp(\alpha_1+\beta x)}{1+\exp(\alpha_1+\beta x)}+(1-p)\frac{\exp(\alpha_2+\beta x)}{1+\exp(\alpha_2+\beta x)}=\frac{\exp(\tilde{\alpha}+\tilde{\beta}x)}{1+\exp(\tilde{\alpha}+\tilde{\beta}x)} \qquad (12)$$

with $\alpha_3=\ln\left(\frac{-pe^{a_1}-e^{a_2+a_1}-(1-p)e^{a_2}}{(p-1)e^{a_1}-e^{a_2}p-1}\right)$ and $\tilde{\beta}=\frac{\ln(-t)-\ln(-w)}{x}$ where *t* and *w* are described in Appendix A. When $\alpha_1=\alpha_2$ we get $\tilde{\beta}=\beta$, see Appendix A. Otherwise, the slope parameters will not coincide. Furthermore, the slope of the "joint" model for both latent classes depends on the value of the independent variable *x*. As a consequence, no joint model exists in which the slope parameter is the same in the two latent classes and, furthermore, no model exists with a uniform slope parameter. These properties are illustrated in Figures 2a and 2b.

FIGURE 2a and 2b HERE

In Figure 2a the two dotted lines represent the predicted probabilities associated with a slope parameter for the independent variable of 1. The model with higher probabilities has a constant term of 2 whereas the model with the lower probability has a constant term of 0. We assume that the

[1] In practice we do not observe the actual probabilities but rather binary 0/1 values of the dependent variable. However, since we are concerned with population level values and not individual estimates we use probabilities, i.e., expected values.

population is split with 50 percent of the observations in each model. This distribution yields probabilities for a joint model shown in the solid line. We show log-odds estimates rather than probabilities in Figure 2b because the effect of the independent variable is non-linear in probabilities. From Figure 2b we find that the slope of the parameter of the independent variable is constant in the two latent class models but that it varies with *x* in the joint model. Consequently, not only will the slope parameter β be biased if unobserved heterogeneity is present in the binary logit model but there is also no single parameter to estimate. In practice, the empirical estimate of the slope parameter in the standard binary logit model is the average of the different slopes across the values of the independent variable and depends on the distribution of the independent variable.

2.2 Identification

The key challenge when estimating the FMBL model concerns identification of the unobserved heterogeneity component. In this section we show how variation in two dimensions of the data: 1) variation in the dependent and independent variables and 2) variation in the number of panels provide information that identifies the latent class parameters in the FMBL model.

The log-likelihood equations for the FMBL model are

$$\frac{\partial \ln L}{\partial \alpha} = \sum_i y_i \frac{Var(y_i)}{P_i} - (1 - y_i)\frac{Var(y_i)}{1 - P_i} \qquad (13.1)$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_i y_i \frac{x_i Var(y_i)}{P_i} - (1 - y_i)\frac{x_i Var(y_i)}{1 - P_i} \qquad (13.2)$$

$$\frac{\partial \ln L}{\partial \varepsilon} = \sum_i y_i \frac{Var(y_i \mid \Xi = \varepsilon)}{P_i} - (1 - y_i)\frac{Var(y_i \mid \Xi = \varepsilon)}{1 - P_i} \qquad (13.3)$$

$$\frac{\partial \ln L}{\partial p} = \sum_i y_i \frac{P_{0i} - P_{\varepsilon i}}{P_i} - (1 - y_i)\frac{P_{0i} - P_{\varepsilon i}}{P_i}, \qquad (13.4)$$

where $Var(y_i) = pP_{0i}(1 - P_{0i}) + (1 - p)P_{\varepsilon i}(1 - P_{\varepsilon i})$ and $Var(y_i \mid \Xi = \varepsilon) = P_{\varepsilon i}(1 - P_{\varepsilon i})$. From the log-likelihood equations we find that when $\varepsilon = 0 \Leftrightarrow P_{0i} = P_{\varepsilon i}$. This means that whenever $\varepsilon = 0 \Rightarrow \frac{\partial \ln L}{\partial p} = 0; \forall \delta_2$, i.e., when there is no information in the data on the value of *p*, the last Equation (13.4) becomes redundant and it is not possible to identify *p*. In practical terms this situation entails that whenever ε approaches 0, i.e., when there is no unobserved heterogeneity, the likelihood function may behave badly and the FMBL is not identified.

We want to use the empirical variation in the *Y* and *X* variable to identify the unobserved heterogeneity, i.e., the latent class parameters. The amount of variation in *Y* and *X* determines whether or not it is possible to identify the latent classes. The posterior allocation of the latent classes Ξ conditional on *Y* and *X* is defined as

$$P\left(\Xi=\varepsilon \mid Y=y, X=x\right) \equiv \frac{P\left(Y=y \mid X=x, \Xi=\varepsilon\right) p}{\sum_{j=1}^{j=2} P\left(Y=1 \mid X=x, \Xi=\varepsilon_j\right)}=\frac{1}{1+\frac{1-p}{p} \frac{1+\exp(\alpha+\beta x+\varepsilon)}{\exp(y \varepsilon)\left\{1+\exp(\alpha+\beta x)\right\}}}. \qquad (14)$$

Equation (14) shows that the allocation into different classes depends on observed values of *y* and *x*. Similarly to Equation (13.4), when $\varepsilon=0$ then Equation (14) reduces to $1-p$ independently of the observed data. With cross-sectional data, observations on *y = 1* renders the information on *y = 0* redundant (once we know *y = 1* we also know that $y \neq 0$), and only variation in *x* can identify the latent classes. To show this formally we differentiate Equation (14) wrt. *x* and equate to 0 to obtain

$$\frac{\partial P\left(\Xi=\varepsilon \mid Y=y, X=x\right)}{\partial x}\Big|_{\beta \neq 0}=\frac{\beta \exp(\alpha+\beta x+y \varepsilon)\left(1-\exp(\varepsilon)\right) p(1-p)}{\left(p\left\{1-\exp(y \varepsilon)\left[1-\exp(\alpha+\beta x)\right]\right\}-1-\exp(\alpha+\beta x+\varepsilon)\left(1-p\right)\right)^2}=0$$

(15)

$$\Rightarrow$$

$$\varepsilon=0, \qquad (16)$$

since the denominator is always defined. Equation (15) shows that whenever *x* varies so does also the posterior probability of observing a latent class membership (except when the latent class effect is 0). Consequently, individuals with different values of the independent variable *x* have different probabilities of belonging to the different latent classes and, in this way, variation in *x* leads to identification of the distribution of the latent classes.

With panel data, i.e., repeated observations of both *Y* and *X*, we have $P(Y_1 = 1) \neq 1 - P(Y_2 = 0)$, where subscript *1* and *2* indexes which wave of the panel the observation belongs to. Consequently, time-varying information on *Y* and *X* will lead to more information about the latent classes. The reason why can be seen from Equations (17.1)-(17.4) below where we show that variation in the dependent variable across panels leads to different posterior probabilities of being allocated to the different latent classes. If these allocation probabilities are identical across panels for varying *Y*, e.g., $Y_1 = 1$ and $Y_2 = 0$, variation in *Y* does not lead to identification of the allocation probabilities. Formally, the allocation probabilities into latent classes across panels can be written as

$$P\left(\Xi = \varepsilon \mid Y_1 = 1, X = x\right) - P\left(\Xi = \varepsilon \mid Y_2 = 0, X = x\right)$$
$$= \frac{1}{1 + \frac{1-p}{p}\frac{1+\exp(\alpha+\beta x+\varepsilon)}{\exp(\varepsilon)\left\{1+\exp(\alpha+\beta x)\right\}}} - \frac{1}{1 + \frac{1-p}{p}\frac{1+\exp(\alpha+\beta x+\varepsilon)}{\left\{1+\exp(\alpha+\beta x)\right\}}} = 0 \qquad (17.1)$$

$$\Rightarrow$$

$$\frac{1+\exp(\alpha+\beta x+\varepsilon)}{\exp(\varepsilon)\left\{1+\exp(\alpha+\beta x)\right\}} = \frac{1+\exp(\alpha+\beta x+\varepsilon)}{\left\{1+\exp(\alpha+\beta x)\right\}} \qquad (17.2)$$

$$\Rightarrow$$

$$\exp(\varepsilon)\left\{1+\exp(\alpha+\beta x)\right\} = \left\{1+\exp(\alpha+\beta x)\right\} \qquad (17.3)$$

$$\Rightarrow$$

$$\varepsilon = 0 \qquad (17.4)$$

Consequently, when *y* varies so does also the posterior probability of belonging to a latent class, except when the latent class membership effect is zero. If both *X* and *Y* vary across panels we have that

$$P\left(\Xi = \varepsilon \mid Y_1 = 1, X = x\right) - P\left(\Xi = \varepsilon \mid Y_2 = 0, X = x'\right) = 0 \qquad (18.1)$$

$$\Leftrightarrow$$

$$\varepsilon = 0 \text{ or } \varepsilon = \ln\left(-\frac{1+\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x')}\right) - \alpha - \beta x\,, \qquad (18.2)$$

and, thus, identification of ε improves when both *Y* and *X* vary. Finally, note that $P\left(\Xi = \varepsilon \mid Y_1 = 1, X = x\right) = P\left(\Xi = \varepsilon \mid Y_2 = 1, X = x'\right)$; i.e., observations that only change values in *x* (and not in *y*) across time do not contribute to the identification of the latent classes.

We may summarize these findings in the following proposition

If $P(Y=1 \mid x, \varepsilon) = \frac{\exp(\alpha+\beta x+\varepsilon)}{1+\exp(\alpha+\beta x+\varepsilon)}, \alpha, \beta$ known, $P\left(\Xi = \varepsilon \mid Y = y, X = x\right) = P\left(\Xi = \varepsilon \mid Y = y', X = x'\right)$,
$y \neq y'$ with $P(Y = y) \neq 1 - P(Y = y')$ or $\mathrm{x} \neq \mathrm{x}' \Rightarrow \varepsilon \mid x, x', y, y' = 0$.

Proof: See Appendix B. The proposition states that if two different posterior probabilities are equal for different values of *x* (the case of cross-sectional data) or for *y* and or *x* (the case of panel data) the distribution of the latent classes is degenerate, at least for the observed data used in the comparison. Hence, data is non-informative with respect to the distribution of the latent classes. And vice versa: if the posterior probabilities differ for different observed (non-redundant) parts of the data this data is informative on the distribution of the latent classes.

In summary, this section has shown that the finite mixture binary logit (FMBL) model is not identified when the effect of the latent classes is 0 or, in other words, when there is no unobserved heterogeneity. We have furthermore shown that variation in *Y* and *X* in cross-sectional and panel data leads to identification of the FMBL model.

3. Simulation Study

3.1 Simulation Results

We run a series of simulations to analyze identification in the FMBL model. Our principal objective is to evaluate the performance of the FMBL with respect to reducing bias in the estimates of the regression coefficient β of the continuous independent variable *x* relative to the standard binary logit model. We run 100 simulations with 500 observations, including repeated observations in panels. The simulations offer varying degrees of identification in terms of the number of panels and variation in *x*, as defined by the number of values of *x*. The simulation parameters and the results from the simulation study are shown in Table 4.

TABLE 4 HERE

Table 4 shows that in a FMBL model with two latent classes, a continuous *x* variable (with an infinite number of values), and five panels estimates of the model parameters α , β , ε , and δ are close to the true values and have small Root Mean Square Errors (RMSE). The table also shows that, in the case of only one panel (i.e., cross-sectional data) and two values of *x*, the FMBL model performs relatively poorly, but nevertheless estimates the regression coefficient β with much less bias than the standard binary logit model in the sense that the estimate is closer to the true value. The intermediate cases show how bias and RMSE in the FMBL model increases model when we use shorter panels. It is also noteworthy that in all cross-sectional simulations the bias of the FMBL model remains at approximately the same level.

3.1 The FMBL Model with Fixed Latent Class Weights

The simulation with only one panel and two values of *x* shows that the FMBL model with two latent classes estimates β more precisely than the standard binary logit model. Consequently, although the FMBL model is weakly identified it still outperforms the standard binary logit model in terms of the precision of the estimate of β . However, we may want to reduce the bias in the FMBL model further by reducing the number of parameters to be estimated. The latent class parameters are the worst identified parameters in the FMBL model. As a consequence, in empirical applications with cross-sectional data it may be difficult to obtain accurate estimates of the latent class parameters. As a means of improving identification of the FMBL model, we propose to fix the parameter for the weight of one of the latent classes (the transformed probabilities of the latent classes, δ_2 , see Equation 8.1 and 8.2) to improve identification. By doing so, we impose the restriction on the unobserved part of the model that we know the proportion of observations that belong to one of the latent classes. This restriction leads to better identification of the FMBL model since it reduces the number of parameters to be estimated. In the next section we, first present the improvements in precision gained from fixing δ_2 and, second, we motivate why fixing the latent class weight is preferable to fixing other parameters in the model.

Table 4 also shows simulation results for the FMBL model with fixed latent class weights (we refer to this model as FMBLfw).[2] Most importantly, in the weakly identified case with one panel and only two values of *x* the FMBLfw model exhibits considerably less bias in the estimate of β than the standard binary logit model. Consequently, in applied research it may be preferable to estimate the FMBL model with a fixed latent class weight rather than to estimate the standard binary logit model. Not surprisingly, when we have multiple panels and values of *x* the FMBL model is less biased than the FMBLfw because the latent class parameters are well-identified and capture the unobserved heterogeneity in the data. An important question when estimating the FMBLfw model concerns how to find the optimal value for the fixed latent class weight δ_2. Rather than setting an arbitrary value one can use out a grid search to find the optimal weight parameter. For example, one could use grid values of δ_2 corresponding to the weights 0.05, 0.15, 0.25, …, 0.95 to find the value of δ_2 that maximizes the log-likelihood of the FMBLfw model. Even though one would need 10 estimations of the FMBLfw model to carry out a grid search this approach is often faster than estimating one FMBL model.[3] From our simulations we have found that the grid search in the FMBLfw model yields a model fit for the FMBLfw which is similar to that of the FMBL model.

There are several reasons why it is preferable to fix the latent class weight rather than other parameters in the model. First, in the likelihood equations, Equation 13.1 to 13.4, we show that the equation for the latent class weight becomes redundant when the latent class effect approaches 0. Accordingly, for some values of the other parameters there is no information on how to choose a particular value of δ_2. Second, fixing the latent class weights does not have much substantive effect on the other parameters. To illustrate this point, we carry out a principal component analysis (PCA) of the estimates in the simulations shown in Table 4. The eigenvalues and eigenvectors of the estimated parameters in the simulations are shown in Table 5.

---

[2] One might also pursue a profile likelihood approach (see Murphy and Van Der Vaart 2000) in which one iterates between maximizing the likelihood with fixed weights and fixing the remaining parameters while estimating the weights. We have used this approach in the simulations but it did not change any of our results.

[3] We have written GAUSS and R programs which implement this grid search routine in the FMBL and the FMBLfw models. These programs are available from the authors upon request.

TABLE 5 HERE

The sum of the eigenvalues is proportional to the variability of the parameters in the different simulations. The greater the variability, the larger the total sum of the eigenvalues. Moreover, the size of each eigenvalue reflects the proportion of the variability in the estimates associated with this eigenvalue. The eigenvector or factor loadings associated with this eigenvalue indicates which parameters contribute to the overall variability associated with this eigenvalue. By comparing the two top panels of Table 5, representing PCA of the simulations on panel data, with the three lower panels, representing PCA on cross-sectional data, we find that the sum of the eigenvalues are much lower in the simulations based on panel data than the eigenvalues in the simulations based on cross-sectional data. This fact reflects the higher accuracy of panel data estimation compared to cross-section estimation.

The first and largest eigenvalue in all simulations corresponds to an eigenvector with high loadings on the constant term, α , and especially on the effect of the latent class, ε . Accordingly, a large part of the RMSE bias in these two parameters is due to the fact that they are correlated. The second-largest eigenvalue, which is of considerable relative size in the cross-sectional simulations, pertains to an eigenvector with a high loading on the weight of the latent class, δ_2. This result suggests that a large part of the RMSE associated with this parameter is uncorrelated with the other parameters or, in other words, that in cross-sectional data δ_2 can take a wide range of values without affecting the other parameters in the model. Consequently, when identification is weak in an empirical application, it is possible to fix δ_2 without inducing much bias in the other model parameters, and especially in β .

FIGURE 3a + 3b HERE

In the simulations in Table 4 we have used a fixed sample size of 500 observations. Obviously, by increasing our sample size we improve identification. To investigate how sensitive our simulation results are to sample size we have run a number of extra simulations with different sample sizes but kept the number of simulations for each sample size at 100. Figure 3a and 3b plot bias and RMSE in a panel model with one continuous *x* variable, five panels, and with increasing sample sizes. It is evident from the figures that both bias and RMSE decrease substantially with increasing sample size both for the FMBL and the FMBLfw models. In no case, not even with small sample sizes, does the bias for the FMBL and the FMBLfw exceed that of the standard binary logit model. Hence, our results suggest that it is feasible to estimate the FMBL model when one has panel data and rich variation in the independent variables. However, from inspecting the RMSE it seems that with very small samples (less than 400 observations) the FMBL is rather unstable. In such cases the FMBLfw may be preferable since we do not induce much extra bias but obtain better precision compared to the FMBL model in which all parameters are estimated. The FMBLfw outperforms the standard binary logit model in every case both in terms of bias and RMSE.

FIGURE 4a + 4b HERE

The scenario is somewhat different in situations with cross-sectional data and limited variation in the independent variables. Figure 4a and 4b show bias and RMSE in a cross-sectional simulation with a single *x* variable with only two values. Here, it is evident that the FMBL model exhibits considerable bias and large RMSE. In fact, for small sample sizes the FMBL model exhibits as much bias as the standard binary logit model and a very large RMSE. For larger sample sizes the FMBL model still has a much larger RMSE than the standard binary logit model. These results suggest that the FMBL model does not perform very well in situations with cross-sectional data and limited variation in the independent variables. By contrast, the FMBLfw performs better in this situation and exhibits much less bias than the standard binary logit model for all sample sizes and also smaller RMSE, at least for sample sizes larger than 200 observations. This result suggests that that the FMBLfw might be useful when identification of the FMBL fails or is weak.

4. Empirical Example – Public Support for Redistribution

In this section we present an empirical illustration of the FMBL and FMBLfw models. We analyze data from the Canadian "Equality, Security, and Community" (ECS) survey, a two-wave panel survey conducted in 2000/2001 (wave 1) and 2002/2003 (wave 2) (see ECS Technical Documentation 1999; Jæger 2006). The ECS includes several subsamples, but we use the National Probability Sample which is representative of adult Canadians. The sample size is around 2,000.

The ECS includes a wide range of binary attitudinal items. In this application we focus on an indicator of whether or not respondents support income redistribution. Public support for income redistribution, and determinants of support for redistribution, has been studied extensively in political sociology (e.g., Edlund 1999; Brooks and Manza 2006; Kenworthy and McCall 2008). The respondents in the ECS were asked: "The government must do more to reduce the income gap between rich and poor Canadians". Respondents answered this question by stating that they either agreed or did not agree with the statement. Table 6 shows that in both waves of the ECS the majority of respondents agree with the statement.

TABLE 6 HERE

We include a number of independent variables in the analysis. First, we include gross personal income in Canadian dollars, here coded into income deciles. Second, we include educational level. Educational level is measured in the ECS survey by ten ordered categories: 1 = "No schooling", 2 = "Some elementary schooling", 3 = "Completed elementary school", 4 = "Completed secondary/high school", 5 = "Some technical, community college", 6 = "Completed technical, community college", 7 = "Some university", 8 = "Bachelor's Degree", 9 = "Master's Degree", and 10 = "Professional degree or doctorate"). Third, we control for the size of the residential are in which the respondent lives. The available categories are 1 = "small town", 2 = "Census Agglomeration", and 3 = "Census Metropolitan Area". Fourth, we control for sector of employment with a dummy variable for being employed in the public sector. Finally, we control for gender (with a dummy variable for males) and age in years.

We run three types of models. First, we estimate standard binary logit, FMBL, and FMBLfw models using only the first wave of the ECS data.[4] Second, we estimate the same models using both waves of the ECS data. Using only the first wave of the data allows us to analyze how the different models behave when we use cross-sectional data. We then compare these results with the more accurate results from the models that use both waves. In the cross-sectional case we expect the standard logit model to be biased, the FMBL to be unstable, and the FMBLfw to be more reliable than both the standard binary logit and the FMBL models.

TABLE 7 HERE

[4] We used a grid search to find the optimal value for the fixed latent class weight in the FMBLfw model. The optimal value for the weight is the value that maximizes the log-likelihood of the FMBLfw model.

Table 7 shows the results from the different model specifications. From the table we see that both in the cross-sectional and panel data models the probability of supporting redistribution is negatively affected by higher income, education, urbanization (size of residential area), and being male. By contrast, public sector employment and age are positively related to support for redistribution. However, we also find that the estimated log-odds ratios vary considerably between the different model specifications and between the cross-sectional and panel data estimations. The FMBL model estimated on panel data is our benchmark model since this model uses the panel information in the data and controls for (well-identified) unobserved heterogeneity and, in doing so, produces the most trustworthy results.

As expected, the standard binary logit model is unaffected by whether or not one uses cross-sectional or panel data. In both cases the model greatly underestimates the log-odds ratios compared to the benchmark FMBL panel data model. The log-odds estimates from the FMBL model based on cross-sectional data are very similar to those in the standard logit model and, thus, are also severely biased. The estimated weight parameter in the FMBL model based on cross sectional data is also very large which indicates poor identification. Results from the FMBLfw model based on cross-sectional data also suggest weak identification since the standard errors of most of the parameter estimates are larger than both those in the standard logit and FMBL models based on cross-sectional data.

To evaluate the consistency of the three different cross-sectional models the last three columns in Table 7 report whether the estimated effects of the independent variables from these models differ significantly from those obtained from our benchmark model, the panel data FMBL model. Here, we find that only in the case of the cross-sectional FMBLfw model do all the estimated effects of the independent variables *not* differ significantly from the results obtained from the panel data FMBL model. In the standard binary logit and the FMBL models three out of six effects do not differ from those of the benchmark model. Consequently, although the FMBLfw model is somewhat imprecise when used with cross-sectional data, it yields much more trustworthy results than both the cross-sectional standard binary logit and FMBL models.

5. Conclusions

Unobserved heterogeneity is particularly important in non-linear regression models. Unobserved heterogeneity leads to biased parameter estimates of the effect of independent variables on the outcome and to incorrect inference. Furthermore, unless one has rich panel data it may be difficult to deal effectively with bias from unobserved heterogeneity.

This paper proposes a new approach to dealing with unobserved heterogeneity in the binary logit model which is useful in applied research. Our approach, which also generalizes to situations with other types of limited dependent variables, builds on the finite mixture framework which models the unobserved heterogeneity via latent classes that capture unobserved sub groups in the data. By modeling membership of these latent classes jointly with the probability of experiencing the binary outcome of interest, it is possible to reduce bias from unobserved heterogeneity. We argue that our Finite Mixture Binary Logit (FMBL) approach might be useful in applied research where it is difficult to identify the unobserved heterogeneity, for example when the researcher has only cross-sectional data or short panels. We present simulation evidence which shows that the FMBL model is superior to the standard binary logit model in terms of reducing bias in the estimated effects of independent variables. Furthermore, we suggest that in situations where the FMBL model is poorly identified, for example in situations with cross-sectional data (or short panels) or with limited variability in the dependent and independent variables, it is useful to fix the parameter for the weight of one or more of the latent classes. Fixing the weight of one of the latent classes has little impact on the other parameters in the model but improves identification and precision, especially in comparison with the standard binary logit model. We also propose a grid search method to find the optimal value of the fixed latent class weight. Finally, we provide an empirical illustration of our new approach using Canadian panel data on public support for redistribution and show that our restricted FMBLfw model is superior to both the FMBL and the standard binary logit model when used with cross-sectional data.

Our new approach contributes to the growing awareness about the impact of unobserved heterogeneity in applied research and the limitations of standard regression models. To address these issues the existing literature generally points to more complicated models. However, theoretical identification of these models is not always clear, and their implementation in practice is often cumbersome. This paper shows how a particular class of models, a binary logit model that allows for unobserved heterogeneity, can be identified from different sources of variation in the data. Although our FMBL model is theoretically identified it may be difficult to estimate with cross-sectional data or short panels. To remedy this problem we suggest simplifying the FMBL model by fixing one of the latent class weights. This simplification makes the model easier to estimate but does not have any substantive impact on the precision of the estimated effects of the independent variable on the binary outcome. The simplified approach might then be useful for applied researchers for whom the main objective is to obtain unbiased estimates of the effects of independent variables.

Appendix A: The function *f(.)* in Equation 12.

$$t = \frac{pe^{\alpha_1} + e^{\alpha_2} + e^{\alpha_2+\alpha_1} - e^{\alpha_2} p}{pe^{\alpha_1} - e^{\alpha_2} p - 1 - e^{\alpha_1}}; w = \frac{-pe^{\alpha_1+\beta x} - e^{\beta x+\alpha_2} - e^{2\beta x+\alpha_2+\alpha_1} + e^{\beta x+\alpha_2} p}{-pe^{\alpha_1+\beta x} + e^{\beta x+\alpha_2} p + 1 + e^{\alpha_1+\beta x}}.$$

When $\alpha_1 = \alpha_2 = \alpha$ we get $t = -e^{\alpha}$ and $w = -e^{a+\beta x}$ and $\tilde{\beta} = \frac{\ln\left(e^{\alpha}\right) + \ln\left(e^{\alpha+\beta x}\right)}{x} = \beta$.

Appendix B: Proof of proposition.

$Proof$: let $e^{\Lambda} = \exp(\alpha + \beta x), e^{\Lambda'} = \exp(\alpha + \beta x')$. Then: $\frac{\frac{\left(e^{\Lambda} e^{\varepsilon}\right)^{y} P(\Xi=\varepsilon)}{1+e^{\Lambda}e^{\varepsilon}}}{\sum_{j=1}^{j=2} \frac{\left(e^{\Lambda} e^{\varepsilon_j}\right)^{y} P(\Xi=\varepsilon_j)}{1+1+e^{\Lambda}e^{\varepsilon_j}}} = \frac{\frac{\left(e^{\Lambda'} e^{\varepsilon}\right)^{y'} P(\Xi=\varepsilon)}{1+e^{\Lambda'}e^{\varepsilon}}}{\sum_{j=1}^{j=2} \frac{\left(e^{\Lambda'} e^{\varepsilon_j}\right)^{y'} P(\Xi=\varepsilon_j)}{1+e^{\Lambda'}e^{\varepsilon_j}}} \Rightarrow$

$$e^{(y'-y)\varepsilon}\left(1 + e^{\Lambda}e^{\varepsilon} + e^{\Lambda'} + e^{\Lambda}e^{\Lambda'}e^{\varepsilon}\right) = 1 + e^{\Lambda} + e^{\Lambda'}e^{\varepsilon} + e^{\Lambda}e^{\Lambda'}e^{\varepsilon} \qquad (*)$$

If $y = y', \ \Lambda \neq \Lambda'$ (cross-sectional data) (*) reduces to: $e^{\Lambda}e^{\varepsilon} + e^{\Lambda'} = e^{\Lambda'}e^{\varepsilon} + e^{\Lambda} \Rightarrow e^{\varepsilon} = 1 \Rightarrow \varepsilon = 0.$
If $y = 0, y' = 1,$ (panel data, the case $y = 1, y' = 0$ being similar) from (*) we get:
$\left(e^{\varepsilon}\right)^2 \left\{e^{\Lambda}e^{\Lambda'}\right\} + e^{\varepsilon}\left\{1 + e^{\Lambda} - e^{\Lambda}e^{\Lambda'}\right\} - \left\{e^{\Lambda} + 1\right\} = 0$ (**). This is a second order polynomial in e^{ε} with roots $-\frac{1+e^{\Lambda}e^{\Lambda'}}{e^{\Lambda}e^{\Lambda'}}$ and 1. *As* the first root is negative we can discard this solution and is only left with $e^{\varepsilon} = 1 \Rightarrow \varepsilon = 0.$

References

Abramson, Charles, Rick L. Andrews, Imran S. Currim, and Morgan Jones 2000. "Parameter Bias from Unobserved Effects in Multinomial Logit Model of Consumer Choice." *Journal of Marketing Research* 37:410-426.

Bearse, Peter, Jos Canals-Cerd, and Paul Rilstone 2007. "Efficient Semiparametric Estimation of Duration Models with Unobserved Heterogeneity." *Econometric Theory* 23:281-308.

Bretagnolle, J., and C. Huber-Carol 1988. "Effects of Omitting Covariates in Cox's Models for Survival Data." *Scandinavian Journal of Statistics* 15:125-138.

Brooks, Clem, and Jeff Manza 2006. "Social Policy Responsiveness in Developed Democracies." *American Sociological Review* 71:474-494.

Davies, Richard B. 1993. "Nonparametric control for residual heterogeneity in modelling recurrent behaviour." *Computational Statistics & Data Analysis* 16:143-160.

Dayton, Chauncey M. 1999. *Latent Class Scaling Analysis*. Thousand Oaks, CA: Sage.

Edlund, Jonas 1999. "Trust in government and welfare regimes: Attitudes to redistribution and financial cheating in the USA and Norway." *European Journal of Political Research* 35:341-370.

ESC Technical Documentation 1999. *Equality, Security and Community (ESC) Wave 1, Data Collection. Technical Documentation.* Toronto: Institute of Social Research, York University.

Goodmann, Leo A. 1974. "Exploratory latent structure analysis using both identifiable and unidentifiable models." *Biometrika* 61:215–231.

Ejrnæs, Mette, and Anders Holm 2006. "Comparing Fixed Effect and Covariance Structure Estimators." *Sociological Methods and Research* 35: 61-83.

Halaby, Charles N. 2004. "Panel Model in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507-544.

Heckman, James J., and Burton Singer 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52:271-320.

Holm, Anders 2002. "The Effect of Training on Search Durations; A Random Effects Approach." *Labour Economics* 9:433-450.

Jæger, Mads Meier 2006. "What makes people support public responsibility for welfare provision: Self-interest or political ideology?" *Acta Sociologica* 49:321-338.

Jæger, Mads, and Anders Holm 2007. “Does parents’ economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime?” *Social Science Research* 36:719-744.

Kenworthy, Lane, and Leslie McCall 2008. “Inequality, Public Opinion, and Redistribution.” *Socio-Economic Review* 6: 35-68.

Lindsay, Bruce G. 1983a. “The geometry of mixture likelihoods: a general theory.” *Annals of Statistics* 11:86-94.

Lindsay, Bruce G. 1983b. “The geometry of mixture likelihoods, Part II: the exponential family.” *Annals of Statistics* 11:783-792.

McCutcheon, Allan L. 1987. *Latent class analysis*. Thousand Oaks, CA: Sage Publications.

Murphy, S. A., and A. W. Van Der Vaart 2000. “On Profile Likelihood.” *Journal of the American Statistical Association* 95:449-465.

Vermunt Jeroen K. 1997. *Log-linear models for event history histories*. Thousand Oaks, CA: Sage Publications.

Acknowledgements

We gratefully acknowledge comments from participants at the seminar in applied statistics held at the University of Aarhus, the Advanced Quantitative Sociologists group at the University of Copenhagen, participants at the RC33 meeting in Naples, 2008 (especially Vernon Gayle and Martin Speiss), and Søren Risbjerg Thomsen and Jan Høgelund.

Table 1. Stylized Data.

	Y = 1	*Y = 0*
X = 0	52	148
X = 1	100	100
X = 2	148	52
X = 3	188	12

Table 2. Stylized Data Stratified by Sub Group

	Group 1		Group 2	
	Y = 1	*Y = 0*	*Y = 1*	*Y = 0*
X = 0	50	50	2	98
X = 1	88	12	12	88
X = 2	98	2	50	50
X = 3	100	0	88	12

Table 3. Parameter Estimates from the Binary Logit and FMBL models

Model	α	β	ε	$P(\Xi = \varepsilon)$	*Log-Likelihood*
Binary logit, no *x*	*0.447*	-	-	-	*- 53499.84*
Binary logit with *x*	*-1.135*	*1.175*	-	-	*- 41479.12*
FMBL, two latent classes	*0.013*	*2.026*	*-4.078*	*0.500*	*- 41324.56*

Table 4. Simulation Results. Bias and RMSE

		Standard Binary logit		FMBL, two latent classes		FMBLfw	
Setup	True value	Bias	RMSE	Bias	RMSE	Bias	RMSE
Five Panels, Values of x: ∞	α=-0.3	-0.636	1.962	0.028	0.011	0.238	0.282
	β=0.4	-0.278	0.318	-0.003	0.013	-0.075	0.033
	ε=0.5	-	-	0.101	0.005	-0.057	0.005
	δ_2=-0.15	-	-	0.060	0.021	-	-
Two Panels, Values of x: 2	α=-0.3	-0.631	1.928	0.071	0.062	0.441	1.005
	β=0.4	-0.260	0.275	0.013	0.018	-0.073	0.032
	ε=0.5	-	-	-0.117	0.039	-2.443	0.196
	δ_2=-0.15	-	-	0.036	0.020	-	-
One Panel, Values of x: 2	α=-0.3	-0.632	1.942	1.221	17.681	1.489	18.663
	β=0.4	-0.263	0.289	0.153	3.186	0.052	0.140
	ε=0.5	-	-	-13.029	15.071	-17.802	15.333
	δ_2=-0.15	-	-	-1.556	27.658	-	-
One Panel, Values of x: 4	α=-0.3	-0.634	1.959	1.049	13.098	0.403	8.486
	β=0.4	-0.274	0.328	0.092	1.449	-0.141	0.192
	ε=0.5	-	-	-11.077	10.914	-0.773	8.352
	δ_2=-0.15	-	-	-1.242	19.597	-	-
One Panel, Values of x: ∞	α=-0.3	-0.633	1.955	1.289	18.455	0.684	11.302
	β=0.4	-0.277	0.347	0.112	1.903	-0.117	0.199
	ε=0.5	-	-	-14.187	4.913	-4.947	10.221
	δ_2=-0.15	-	-	-1.327	24.266	-	-

Note: δ_2 fixed to 0. Number of observations = 500. Bias is the average deviation between the estimated values of the parameters and the true values.

Table 5. Principal Component Analysis of Simulation Results.

	PC1	PC2	PC3	PC4
Five panels, two values of x				
Eigenvalues	1.386	0.112	0.075	0.035
α	0.448	-0.417	0.257	-0.748
β	0.001	0.164	-0.901	-0.401
ε	-0.892	-0.155	0.148	-0.398
δ_2	-0.055	-0.881	-0.316	0.350
Two panels, two values of x				
Eigenvalues	1.683	0.422	0.145	0.078
α	0.471	-0.461	0.421	-0.623
β	-0.012	0.045	-0.808	-0.588
ε	-0.879	-0.183	0.260	-0.354
δ_2	-0.066	-0.867	-0.320	0.375
One panel, two values of x				
Eigenvalues	5.481	4.146	1.530	1.018
α	0.383	-0.328	0.852	-0.143
β	0.017	0.036	-0.159	-0.987
ε	-0.915	-0.262	0.298	-0.073
δ_2	0.126	-0.907	-0.401	0.034
One panel, two values of x				
Eigenvalues	5.448	3.801	1.866	1.233
α	0.408	-0.241	0.620	0.625
β	-0.025	0.030	-0.696	0.717
ε	-0.902	-0.256	0.260	0.232
δ_2	0.141	-0.936	-0.253	-0.202
One panel, two values of x				
Eigenvalues	6.182	4.559	2.146	1.282
α	0.365	-0.254	0.807	0.388
β	-0.097	0.069	-0.376	0.919
ε	-0.926	-0.113	0.356	0.056
δ_2	0.005	-0.958	-0.283	-0.043

Table 6. Summary Statistics for Variables in ECS Survey

Year	2000/2001		2002/2003	
Dependent variable:	Mean	SD	Mean	SD
Q: "The government must do more to reduce the income gap between rich and poor Canadians …" (percent in agreement)	0.804	0.396	0.753	0.431
Income (deciles)	5.771	2.526	5.807	2.633
Educational level (ten groups)	5.520	2.094	5.626	2.104
Size of residential area[a]	2.220	0.903	2.220	0.903
Public sector employee	0.304	0.460	0.223	0.416
Gender (= male)[a]	0.427	0.495	0.427	0.495
Age[a]	45.944	15.756	45.944	15.756

Note: [a] variable appears only in 2000/2001 survey.

Table 7. Results from Different Specifications of the Model for Support for Redistribution. Log-Odds Estimates with Standard Errors in Parenthesis

	Cross-sectional (wave 1)			Panel (waves 1 and 2)			Estimates from cross-sectional models within 95 percent CI of the two panel FMBL model		
	Binary Logit	FMBL	FMBL fw	Binary Logit	FMBL	FMBL fw	Binary Logit	FMBL	FMBL Fw
Constant	2.349	2.408	0.060	2.464	1.062	1.239	-	-	-
	(0.299)	(1.488)	(0.523)	(0.216)	(0.544)	(0.451)			
Income/10	-1.188	-0.882	-6.102	-1.315	-1.949	-1.947	No	No	Yes
	(0.237)	(0.254)	(2.280)	(0.176)	(0.328)	(0.328)			
Educational level/10	-0.469	-0.614	-1.394	-0.679	-1.351	-1.416	No	No	Yes
	(0.294)	(0.314)	(1.154)	(0.215)	(0.445)	(0.438)			
Size of residential area	-0.170	-0.148	-0.644	-0.146	-0.266	-0.253	Yes	Yes	Yes
	(0.066)	(0.071)	(0.367)	(0.048)	(0.107)	(0.102)			
Public sector employee	0.399	0.100	0.479	0.378	0.731	0.709	No	No	Yes
	(0.142)	(0.147)	(0.473)	(0.098)	(0.172)	(0.175)			
Gender (= male)	-0.246	-0.210	-0.946	-0.199	-0.385	-0.391	Yes	Yes	Yes
	(0.116)	(0.125)	(0.556)	(0.086)	(0.190)	(0.193)			
Age/100	0.500	0.625	2.955	0.528	0.771	0.671	Yes	Yes	Yes
	(0.379)	(0.416)	(1.501)	(0.280)	(0.624)	(0.600)			
ε	-	0.115	9.233	-	3.996	3.987	-	-	-
		(3.620)	(2.167)		(0.172)	(0.174)			
δ	-	0.214	-	-	-1.195	-	-	-	-
		(21.517)			(0.147)				
BIC									
Log-Likelihood	-999.67	-902.94	-898.13	1901.91	-1725.38	-1725.60			

Note: Number of observations = 2,112.

Figure 1. Observed and predicted probabilities for *Y* = *1*.

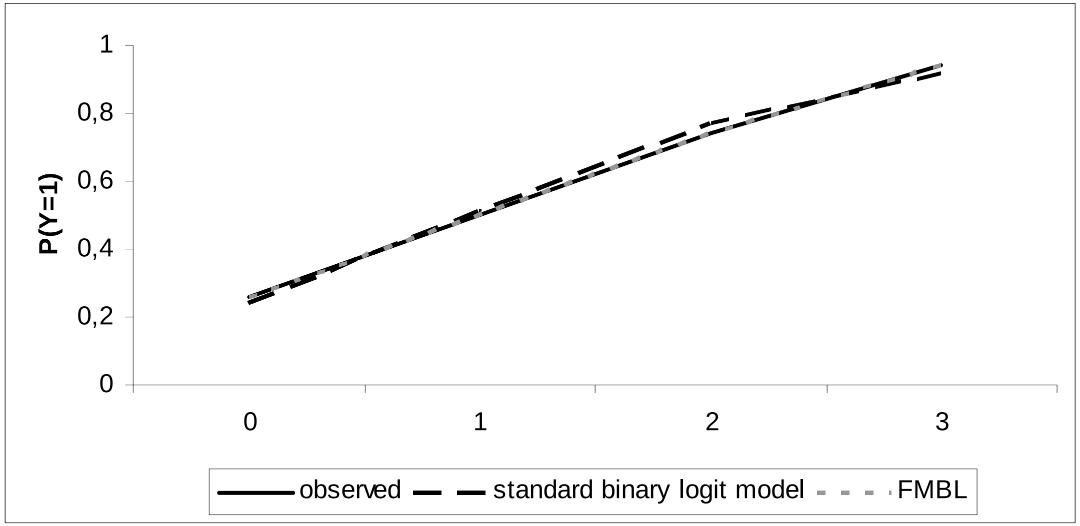

Figure 2a. Probabilities for to logit models with equal slope and mixed probabilities.

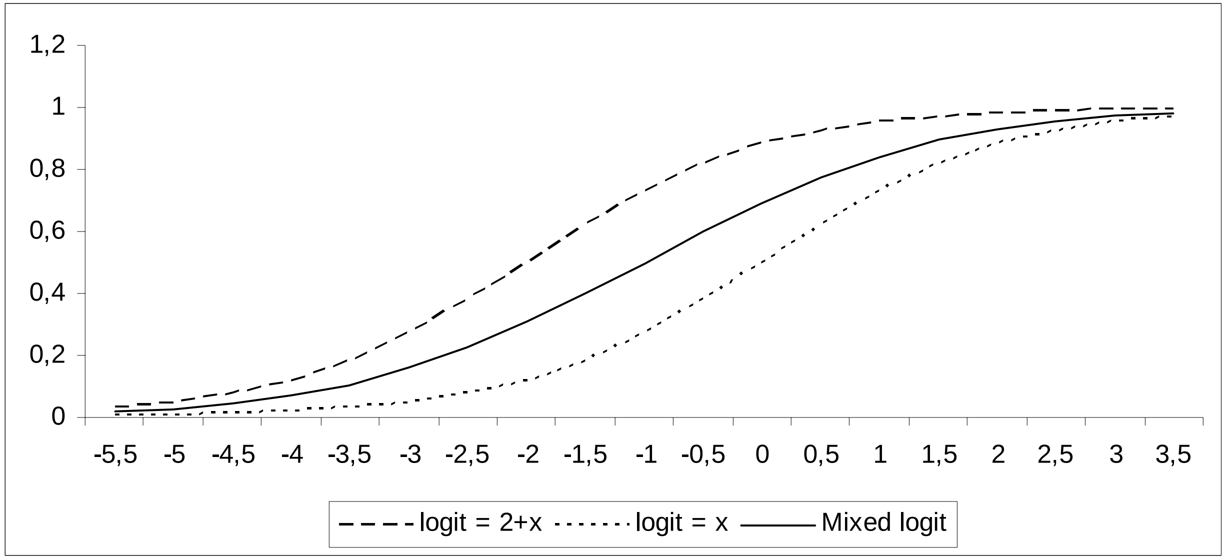

Figure 2b Log-odds for to logit models with equal slope and mixed probabilities.

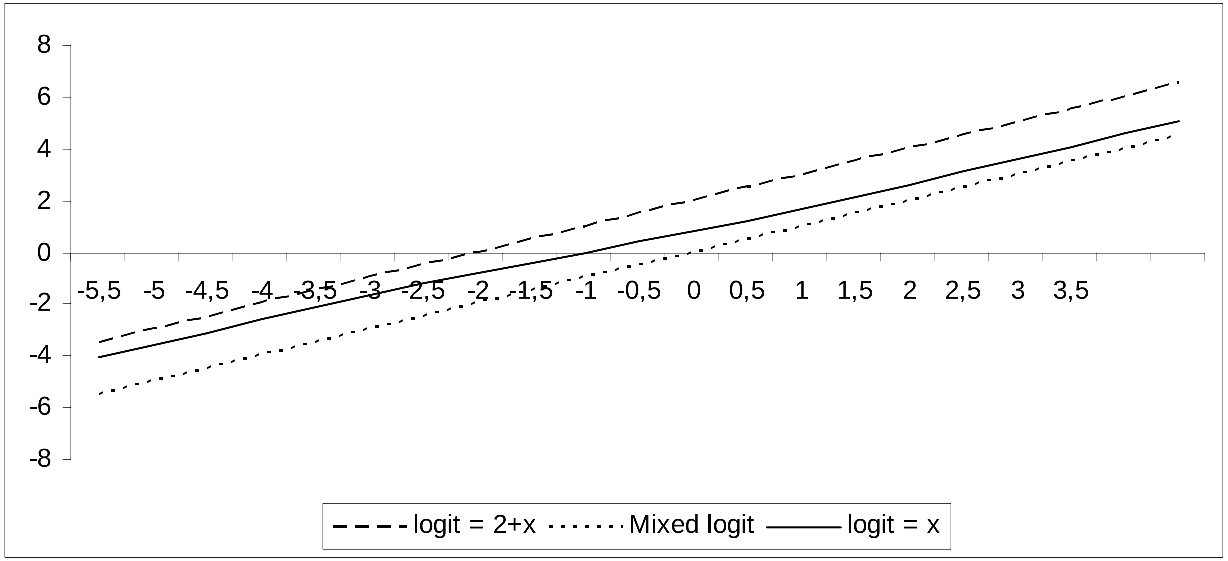

Fig. 3a. Bias in Estimates of β, Panel Data Simulations

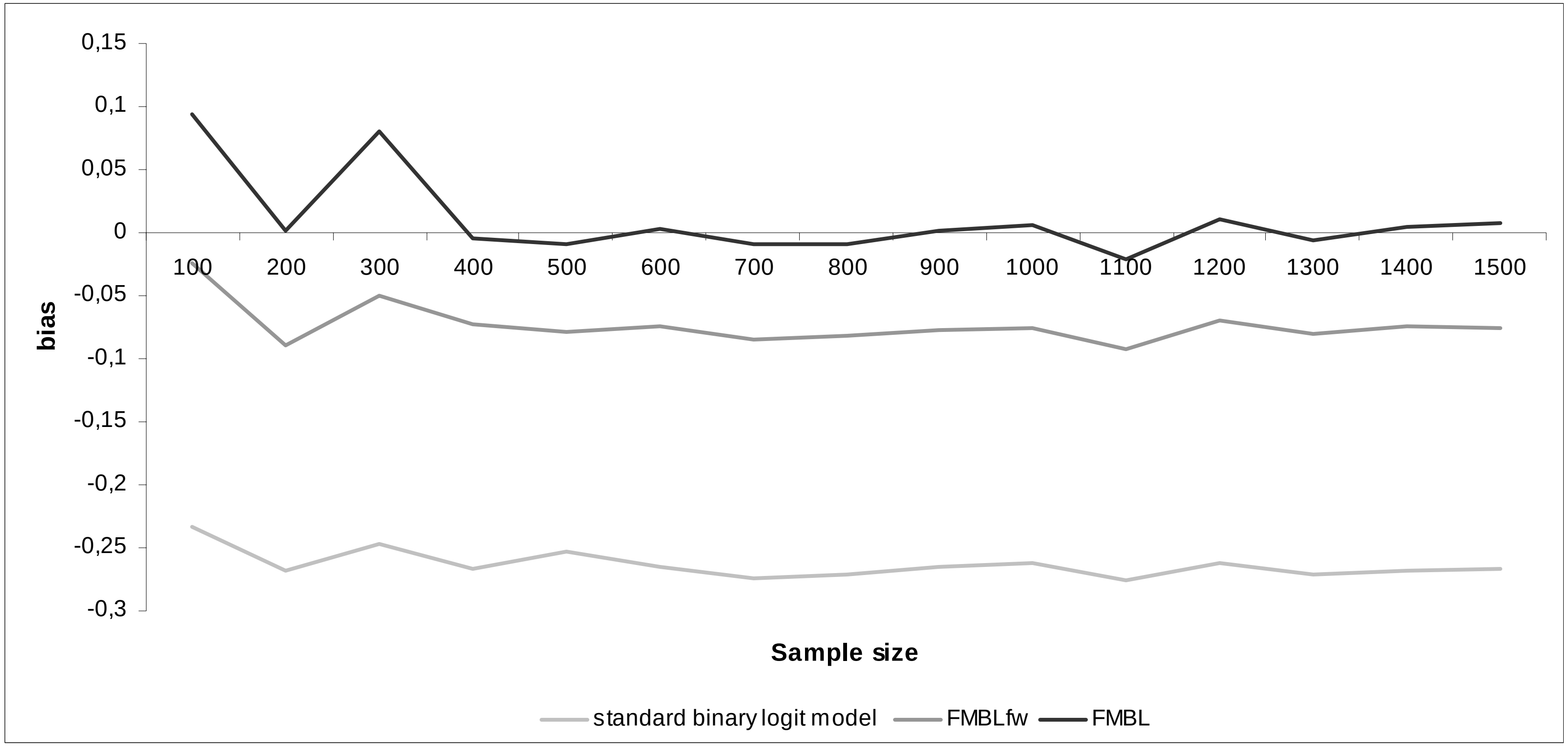

Fig. 3b. RMSE in Estimates of β, Panel Data Simulations

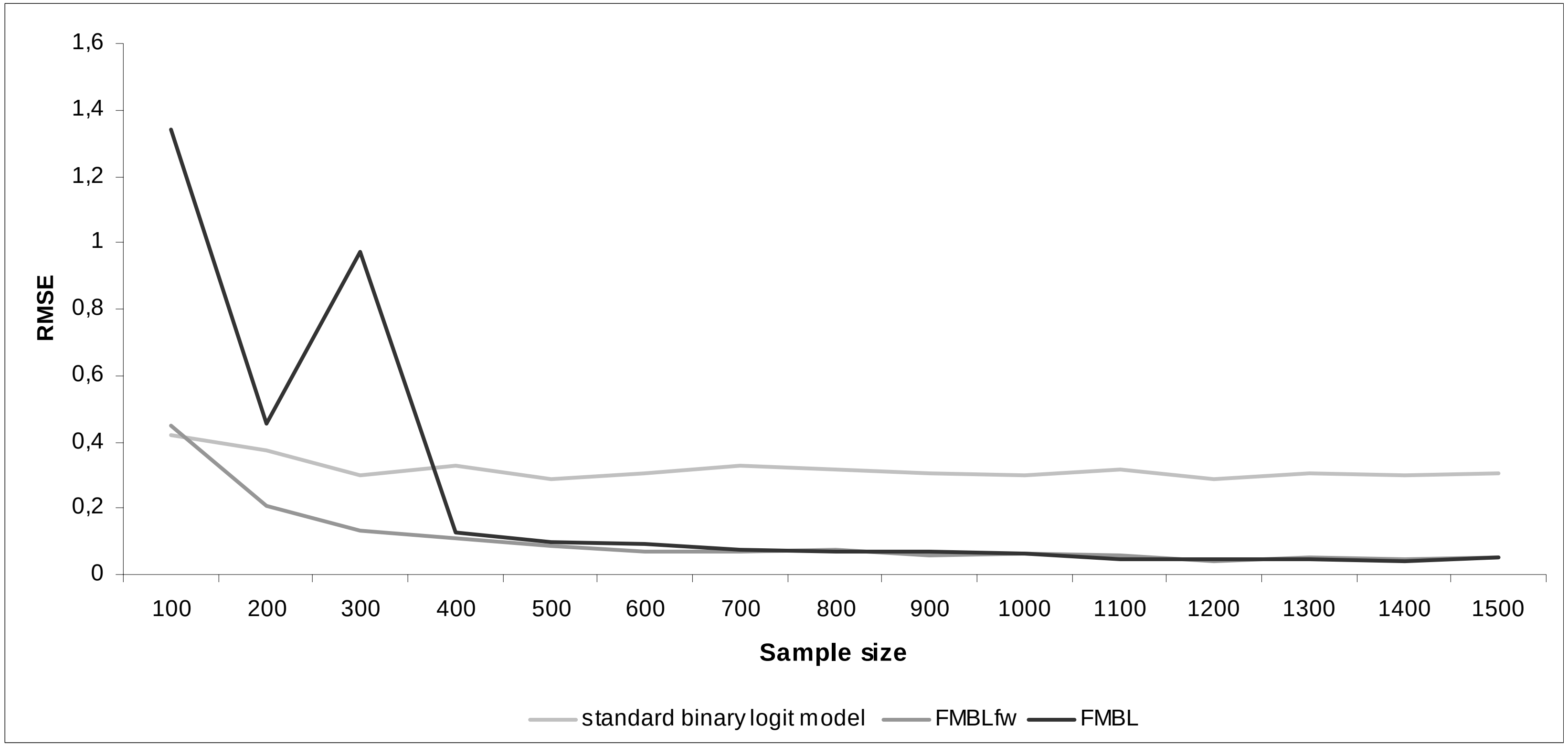

Fig. 4a. Bias in Estimates of β, Cross-Sectional Simulations

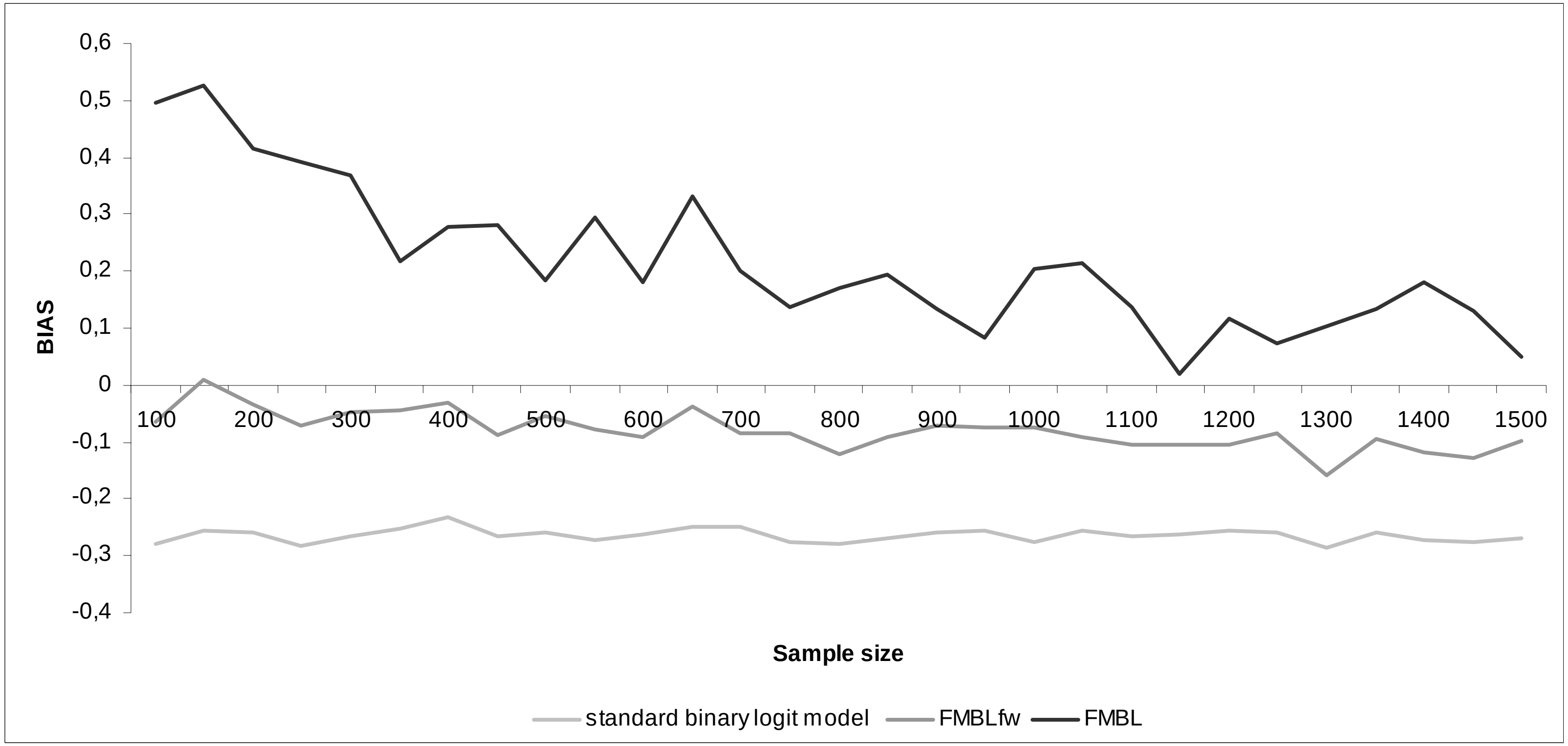

Figure 4b. RMSE in Estimates of β, Cross-Sectional Simulations