Latent variable modelling with non-ignorable item non-response: multigroup response propensity models for cross-national analysis

Jouni Kuha, Myrsini Katsikatsou and Irini Moustaki

London School of Economics and Political Science, UK

Summary. When missing data are produced by a non-ignorable non-response mechanism, analysis of the observed data should include a model for the probabilities of responding. We propose such models for non-response in survey questions which are treated as measures of latent constructs and analysed by using latent variable models. The non-response models that we describe include additional latent variables (latent response propensities) which determine the response probabilities. We argue that this model should be specified as flexibly as possible, and we propose models where the response propensity is a categorical variable (a latent response class). This can be combined with any latent variable model for the survey items, and an association between the latent variables measured by the items and the latent response propensities then implies a model with non-ignorable non-response. We consider in particular such models for the analysis of data from cross-national surveys, where the non-response model may also vary across the countries. The models are applied to data on welfare attitudes in 29 countries in the European Social Survey.

Keywords: Cross-national surveys; Latent class model; Latent trait model; Missing data; Non-ignorable non-response; Response propensity

1. Introduction

Missing data are a persistent problem in many kinds of research. They lead to loss of information and potential bias in the conclusions that are drawn from data. The biases may be avoided if the missing data are adequately accounted for in the analysis. How this may be done depends on the nature of the missingness. A crucial distinction is whether the missing data mechanism is missingness at random (MAR) or missingness not at random (MNAR), i.e. whether the probability of missingness does not or does depend on data which are themselves unobserved (Rubin, 1976; Little and Rubin, 2002). MAR is usually *ignorable* for likelihood-based estimation, meaning that estimation can be done without an explicit model for the probabilities of missingness. MNAR, in contrast, is *non-ignorable*, so a model for it does need to be included in the analysis to obtain valid estimates and inference also for the models of main interest. In this paper, we propose and apply a particular class of models which accommodates both ignorable and non-ignorable missingness.

Our models are used for data with multiple observed variables (*items*) which are treated jointly as measures of unobserved (latent) constructs. Such items are commonly used, particularly in survey research and psychological testing, to measure constructs such as attitudes and abilities.

Address for correspondence: Jouni Kuha, Department of Statistics, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK.
E-mail: j.kuha@lse.ac.uk

As an example, we consider survey items on individuals' attitudes towards the recipients of welfare benefits. The data are analysed by using *latent variable models* which represent the measurement of the latent constructs by the items. Substantive research questions typically centre on associations involving the latent variables, e.g. how an individual's attitude depends on covariates such as age and education. The kind of missing data that we focus on is *item non-response* which arises when a respondent fails to answer one or more of the items.

We consider models where the probability of response may depend on the latent variables ('attitudes') which are measured by the survey items. We focus on methods where this idea is implemented by introducing one or more additional latent variables ('response propensities') which determine the response probabilities. The non-response is non-ignorable if these response propensities (or the response probabilities themselves) are associated with the latent attitudes, and ignorable otherwise. Both cases are identified from the observed data under conventional assumptions for latent variable models. The models can be used to study the patterns of missingness and characteristics of non-respondents, to test the hypothesis of ignorable non-response, and to examine how different assumptions about the non-response affect estimated models for the latent attitude variable. Different versions of this idea have been proposed previously by, for example, Knott *et al.* (1990), O'Muircheartaigh and Moustaki (1999), Holman and Glas (2005), Rose *et al.* (2010), Glas *et al.* (2015) and Rose *et al.* (2015); these and related literatures are reviewed in Section 4.

Most of the recent developments of response propensity models have taken place in the context of applications to educational and psychological testing. We transfer and extend these methods to the analysis of general social surveys. In particular, we focus on *cross-national surveys* where the same questions are asked of respondents in several countries. Our application uses data from 29 countries in the European Social Survey (ESS) (`http://www.europeansocialsurvey.org/`; Stoop *et al.* (2010), chapter 3), and this work was motivated by an on-going process of research and development around the ESS. Cross-national surveys are conducted to answer comparative questions, such as how the levels of an attitude or its associations with covariates may vary between countries. These questions can be addressed with *multigroup* latent variable models, treating the country as an additional explanatory variable. However, cross-national data may also introduce the complication of cross-national variation in item non-response, i.e. that its levels and mechanisms may vary across the countries. In this situation, researchers would want to be able to examine the nature and extent of such variation, and to assess how it might affect the comparative conclusions from the multigroup analysis.

In this paper we propose, first, a new general multigroup model which includes a latent response propensity model for non-response. We argue that to be sufficiently flexible for cross-national survey applications this non-response model should itself be a 'non-equivalent' multigroup model, which means that both the levels of response propensity and how the response probabilities depend on this propensity should be allowed to vary between countries. We then propose a specific model of this type which is particularly well suited for providing a flexible specification for the non-response model. In this model, the response propensity is a categorical variable or a 'latent response class'. Finally, we demonstrate the use of the model with an analysis of the ESS data, where a latent response class model is combined with a structural equation model for a continuous latent attitude towards welfare recipients, both depending on a respondent's country and individual characteristics.

Multigroup latent variable models for multivariate items are outlined in Section 2. In Section 3 we then discuss models for non-response and define the latent response propensity formulation of them in general and the latent response class model in particular. Previous literature on latent response propensities and related ideas is reviewed in Section 4. The application to the ESS

data is presented in Section 5. The results suggest that non-response in the survey items is non-ignorable in more than half of the 29 countries, in all of them in the direction that individuals with more positive attitudes towards welfare recipients are less likely to give an answer to the items. However, cross-national conclusions about the attitude are not substantially altered by allowing for the non-ignorable non-response, indicating that an analysis which assumes ignorable non-response is adequate in this example. Concluding comments are given in Section 6.

2. Latent variable models: structural and measurement models

Consider data for *n* units such as survey respondents or subjects in an educational assessment. We shall first specify models for an individual unit, so a respondent subscript is omitted from the notation for now. Let $\mathbf{Y}_C = (Y_1, \ldots, Y_p)'$ denote a vector of *p* observable variables (items) for a single respondent. They are regarded as measures of a latent variable η, representing some construct of interest. We focus on the case of a univariate η because it is sufficient for introducing the key ideas, but the model framework extends in a straightforward way also to cases with multiple latent variables.

Some or all of $\mathbf{Y}_C$ may be missing. Let $\mathbf{Y}$ denote the items which are observed for a unit, and $\mathbf{Y}_{\mathrm{mis}}$ the missing items, where $\mathbf{Y}$ or $\mathbf{Y}_{\mathrm{mis}}$ may be empty. Let $\mathcal{O} \subset \{1, \ldots, p\}$ denote the indices of $\mathbf{Y}$, and let $\mathbf{R} = (R_1, \ldots, R_p)'$ where $R_j = 1$ if Y_j is observed (i.e. $j \in \mathcal{O}$) and $R_j = 0$ if Y_j is not observed. In addition, we may also observe a vector $\mathbf{X}$ of variables which are treated as covariates (explanatory variables) for η rather than as measures of it. Here $\mathbf{X}$ may also include functions of covariates such as interactions and quadratic terms. We assume that all of $\mathbf{X}$ are observed; if they are not, techniques such as multiple imputation, for example, can be used to allow for non-response in $\mathbf{X}$ (see for example Carpenter and Kenward (2013)).

Conditionally on $\mathbf{X}$, the joint distribution of the other variables can be written as

$$p(\mathbf{R}, \mathbf{Y}_C, \eta | \mathbf{X}) = p(\mathbf{R} | \mathbf{Y}_C, \eta, \mathbf{X}) \, p(\mathbf{Y}_C | \eta, \mathbf{X}) \, p(\eta | \mathbf{X}), \tag{1}$$

where $p(\cdot | \cdot)$ denotes a conditional probability or probability density function. We begin in this section by discussing the *structural model* $p(\eta | \mathbf{X})$ and the *measurement model* $p(\mathbf{Y}_C | \eta, \mathbf{X})$. Together they define a latent variable model with covariates, in which the structural model is typically the focus of substantive research questions. In Section 3 we then discuss the *non-response model* $p(\mathbf{R} | \mathbf{Y}_C, \eta, \mathbf{X})$, which describes the probabilities of missingness in the items, and we propose a latent response propensity specification for this model.

Different choices for the structural and measurement models lead to different latent variable models, such as linear factor analysis (structural equation) models, latent class models and item response theory models (see for example Bartholomew *et al.* (2011) for an overview). We take such commonly used models as our starting point, without discussing their characteristics in further detail (e.g. their parametric distributional assumptions and robustness to them; for discussions of this topic, see for example Boomsma and Hoogland (2001) and Wall *et al.* (2015)). In the application in Section 5 we shall use factor analysis models, where both η and $\mathbf{Y}_C$ are treated as continuous variables, and for this reason their formulae are shown below. The model framework that is described here is not, however, specific to this case, but any latent variable models for η may be combined with the non-response models that are described in Section 3.

We consider, in particular, situations where each respondent belongs to one of *G* distinct groups. In cross-national applications the group is the respondent's country, but the same ideas apply also with any other groupings such as regions or ethnic groups. Allowing for the group

in latent variable modelling for cross-group comparisons is known as *multigroup analysis*. Literature on it goes back to at least Jöreskog (1971) for factor analysis and Clogg and Goodman (1984) for latent class models, and a current overview can be found in Kankaraš *et al.* (2011).

The structural model $p(\eta|\mathbf{X})$ is a regression model for the latent variable η given covariates $\mathbf{X}$. In our application, η in group $g = 1, \ldots, G$ is taken to be normally distributed as

$$\eta \sim N(\gamma_0^{(g)} + \gamma_x^{(g)\prime}\mathbf{X}, \psi^{(g)}) \tag{2}$$

where $\gamma_0^{(g)}$, $\gamma_x^{(g)}$ and $\psi^{(g)}$ are parameters. The superscript (g) indicates a parameter which applies in a specific group, and a parameter without a superscript is taken to be equal across the groups.

For the measurement model we assume throughout that $p(\mathbf{Y}_C|\eta, \mathbf{X}) = p(\mathbf{Y}_C|\eta)$ and that its parameters do not depend on the group. This is the assumption of *measurement equivalence* with respect to the groups and the covariates (see for example Millsap (2011)). We also assume that $p(\mathbf{Y}_C|\eta) = \Pi_{j=1}^p p(Y_j|\eta)$, i.e. that the items Y_j are conditionally independent given η. Both of these assumptions can be relaxed, but doing so is separate from the issues of item non-response which are our main focus. The forms of the measurement models for the individual items Y_j depend on the types of the Y_j and η. For a linear factor analysis model used in our application we have, given the continuous η,

$$Y_j|\eta \sim N(\tau_j + \lambda_j\eta, \theta_j) \qquad \text{for } j = 1, \ldots, p, \tag{3}$$

where the parameters τ_j, λ_j and θ_j are the measurement intercept, factor loading and error variance respectively, and the values of these parameters do not vary across the groups when we assume measurement equivalence. Other choices would be used for other types of items, e.g. logistic measurement models for categorical Y_j. To identify the scale of η, model (2)–(3) requires two parameter constraints, e.g. $\gamma_0^{(g)} = 0$ and $\psi^{(g)} = 1$ for $g = 1$.

3. Latent response class models for non-response

3.1. *Missingness mechanisms and response propensity models*
The non-response model $p(\mathbf{R}|\mathbf{Y}_C, \eta, \mathbf{X})$ represents the missing data mechanism. In this section we first discuss some general ideas related to it and then introduce the idea of specifying it in terms of a latent response propensity. The specific (latent response class) non-response models that we propose are then described in Section 3.2 and defined by equations (7)–(9) there.

Missing items $\mathbf{Y}_{\mathrm{mis}}$ for a unit are missing at random if

$$p(\mathbf{R} = \mathbf{r}|\mathbf{Y} = \mathbf{y}, \mathbf{Y}_{\mathrm{mis}}, \eta, \mathbf{X} = \mathbf{x}) = p(\mathbf{R} = \mathbf{r}|\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}) \tag{4}$$

for all possible values of $\mathbf{Y}_{\mathrm{mis}}$, η and the parameters of this distribution, where $\mathbf{r}$, $\mathbf{y}$ and $\mathbf{x}$ denote the observed values of $\mathbf{R}$, $\mathbf{Y}$ and $\mathbf{X}$ (Rubin, 1976). In other words, missingness is MAR if, given the observed values of the non-missing variables, the probability of the observed pattern of non-response does not depend on any of the unobserved variables, which here include both $\mathbf{Y}_{\mathrm{mis}}$ and η. If this does not hold, the items are missing not at random.

We shall not consider models where $\mathbf{R}$ depends directly on the missing $\mathbf{Y}_{\mathrm{mis}}$. This would define a type of MNAR which cannot be checked within our model without further assumptions. Next, consider the role of $\mathbf{Y}$ in equation (4). This is carefully defined to be the most general condition under which ignorability will follow. It needs to hold only for the specific configuration and values of the variables that are observed in the data. In particular, MAR is satisfied if, for every respondent, $\mathbf{R}$ depends only on those items $\mathbf{Y}$ which are observed for that respondent, even if the set of items which make up $\mathbf{Y}$ varies from one respondent to the next. However, this

assumption is so specific that it is difficult to think of it as a general data-generating model for $\mathbf{R}$. To obtain such a model, we would want to condition it only on those items in $\mathbf{Y}_C$ which are observed for every respondent. In many applications this means none of them, because every item is missing for some respondents. Thus MAR for multiple items often means in practice that the missingness should not depend directly on any of the items themselves. We shall consider only non-response models of this kind, where $p(\mathbf{R}|\mathbf{Y}_C, \eta, \mathbf{X}) = p(\mathbf{R}|\eta, \mathbf{X})$ for all values of the variables. This is still MNAR but it becomes MAR if also $p(\mathbf{R}|\eta, \mathbf{X}) = p(\mathbf{R}|\mathbf{X})$, i.e. if, given the covariates, non-response does not depend on the latent η.

As η and $\mathbf{Y}_{\mathrm{mis}}$ are not observed, estimation is based on the distribution of the observed variables only. Starting from equation (1), and with the further assumptions stated so far, this is

$$p(\mathbf{R}, \mathbf{Y}|\mathbf{X}) = \int \, p(\mathbf{R}|\eta, \mathbf{X}) \Big\{ \prod_{j \in \mathcal{O}} p(Y_j|\eta) \Big\} p(\eta|\mathbf{X}) \mathrm{d}\eta \tag{5}$$

where the integral is over the possible values of η. If the missingness is MAR, this simplifies further so that $p(\mathbf{R}|\eta, \mathbf{X}) = p(\mathbf{R}|\mathbf{X})$ moves outside the integral in equation (5). If then the parameters of the non-response model and of the structural and measurement models are also distinct from each other—which we assume throughout—the missingness in $\mathbf{Y}_C$ is *ignorable*. This means that valid likelihood-based estimation and inference of the structural and measurement models can be done based only on $p(\mathbf{Y}|\mathbf{X}) = \int \{\Pi_{j \in \mathcal{O}} p(Y_j|\eta)\} p(\eta|\mathbf{X}) \mathrm{d}\eta$, omitting the non-response model from the likelihood (see Molenberghs and Kenward (2007), chapter 12). If, instead, $\mathbf{R}$ does depend on η, the non-response is non-ignorable and the non-response model does need to be included in the estimation. This also is possible here because an MNAR non-response model is also identified in a latent variable modelling framework, in essence because the distribution of the latent η is identified from the multiple observed variables $(\mathbf{Y}, \mathbf{R})$.

What remains to be done is to specify the form of the non-response model. Here we focus on the family of models which is obtained by introducing one or more additional latent variables ξ such that $p(\mathbf{R}|\eta, \mathbf{X}) = \int p(\mathbf{R}|\xi, \eta, \mathbf{X}) p(\xi|\eta, \mathbf{X}) \mathrm{d}\xi$. We refer to ξ as *latent response propensities*, because they represent unobservable determinants of non-response—e.g. the propensity of an individual to respond or not to respond to survey items. We assume further that the response indicators R_j for different items j are conditionally independent given $(\xi, \eta, \mathbf{X})$, so that the non-response model is of the form

$$p(\mathbf{R}|\eta, \mathbf{X}) = \int \Big\{ \prod_{j=1}^{p} p(R_j|\xi, \eta, \mathbf{X}) \Big\} p(\xi|\eta, \mathbf{X}) \, \mathrm{d}\xi. \tag{6}$$

Here $p(\xi|\eta, \mathbf{X})$ describes how the response propensity ξ depends on attitudes η and covariates $\mathbf{X}$, and model (6) describes how ξ determines the probabilities and associations of the response indicators $\mathbf{R}$. This defines MNAR if either of the distributions on the right-hand side of equation (6) depends on η, i.e. if η and ξ are associated given $\mathbf{X}$ and/or if the R_j depend directly on η. That such non-response is MNAR and thus non-ignorable was pointed out already by Muthén *et al.* (1987). Frangakis and Rubin (1999) and Harel and Schafer (2009) referred to it as 'latent ignorability', because the non-response would be ignorable if the latent variables were observed.

Model (6) can be combined with any specifications for the structural and measurement models, to define the joint model (5). Fig. 1 shows a path diagram which represents such a model. It has two latent variables (ξ, η) which are measured by two sets of observed indicators, $\mathbf{R}$ and the items $\mathbf{Y}$. We limit our discussion to models which further assume that $p(R_j|\xi, \eta, \mathbf{X}) = p(R_j|\xi, \mathbf{X})$ for all items j, so that the association between η and $\mathbf{R}$ is mediated entirely via ξ (in Fig. 1
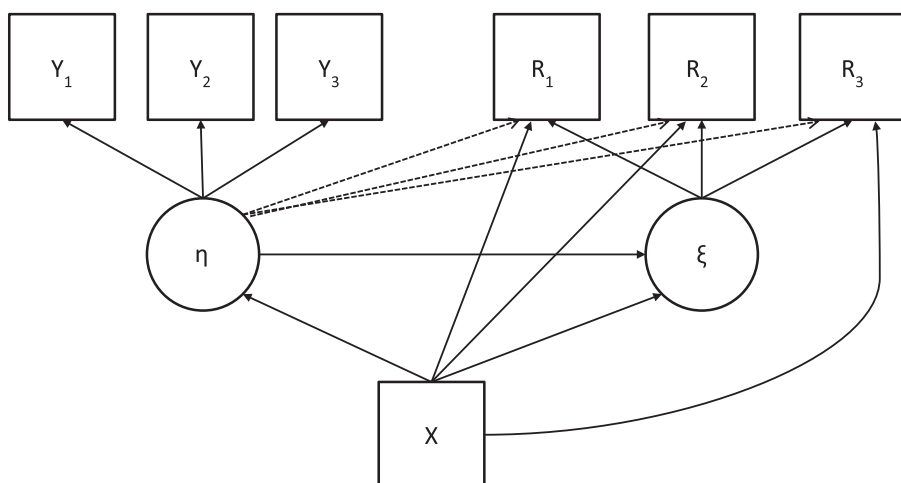
Fig. 1. Path diagram representing the type of joint model that is considered in the paper, i.e. a model for latent variables η measured by items Y_j and conditional on covariates **X**, combined with a model for response indicators R_j given latent response propensities ξ (the associations shown with broken arrows (direct effects of η on R_j) are possible in this framework, but not used in the examples of the paper): in this diagram dependence on **X** includes also dependence on group in multigroup applications, which in the notation of the paper is indicated separately by superscript (*g*)

this means omitting the broken arrows). This has the advantage that the hypothesis of ignorable non-response has the simple form that η and ξ are conditionally independent, i.e. that $p(\xi|\eta, \mathbf{X}) = p(\xi|\mathbf{X})$, but at the cost of the constraint that the non-response is either ignorable for all of the items in $\mathbf{Y}_C$ or for none of them. The models that we focus on are of this form and with ξ specified as a categorical variable, as described next.

3.2. Latent response class models

Modelling the response propensity ξ can be of interest in itself, separately from models for the latent variable η. Often, however, this not so, so that η remains the only subject of substantive research questions whereas ξ is primarily a device which is introduced to define a non-response model. The main purpose of ξ is then to enable the testing of and, if necessary, allowing for non-ignorable non-response in η. This has the implication that the non-response model should be sufficiently richly specified, so that inappropriate constraints in it do not end up distorting conclusions about the models for η. In particular, we would not want to treat the non-response as non-ignorable unnecessarily, if the data would allow us to conclude that it is in fact ignorable. In the context of multigroup analysis which is our particular concern, we would also want to allow for possible differences in the non-response model between different groups such as countries. Such considerations suggest that parsimony and easy interpretability are less important in the non-response model than are richness and flexibility. We may also be indifferent to the exact form of this model, as long as its goals are achieved.

There are three complementary ways of defining a flexible non-response model. The first is to include a rich set of covariates **X**, to increase the chance that η and ξ can be taken to be independent given **X** (this point is discussed by Collins *et al.* (2001), Graham (2003) and Rose *et al.* (2015)). Even if the goal is to estimate the distribution of η given a smaller set $\mathbf{X}^*$ of covariates, it may thus be desirable to fit models first given a larger set $\mathbf{X} = (\mathbf{X}^*, \mathbf{X}^\dagger)$ and then to obtain an estimate of $p(\eta|\mathbf{X}^*)$ by averaging $p(\eta|\mathbf{X})$ over some reference distribution for $\mathbf{X}^\dagger$.

Second, $p(R_j|\xi, \mathbf{X})$ may depend on $\mathbf{X}$ or on the respondent's group. In particular, in cross-national analyses we want to be able to include non-equivalence of this kind by country. If we let all parameters of these models vary by country, this means in effect that we allow each country to have its own separate non-response model.

Third, $p(\mathbf{R}|\mathbf{X})$, obtained by integrating equation (6) over ξ and η, should specify a sufficiently rich dependence structure for the joint distribution of $\mathbf{R}$. For a given $\mathbf{X}$, this is determined by the specification of the latent variables. Since the choice of η should be driven only by substantive considerations and goodness of fit of the model for $\mathbf{Y}_C$, the form of the non-response model is really determined by the specification of ξ. In almost all of the previous literature ξ has been specified as a continuous variable. In the on-line supplementary materials to this paper we propose multigroup extensions of this specification for use with cross-national data, in a form which allows direct comparison with the latent response class models that are proposed below.

Our main focus, however, will be on models where ξ is not continuous but a categorical variable. This is a choice that we can make even when the substantively interesting latent variable η remains continuous (as it is in our examples in Section 5). Suppose that such a ξ has C categories, which we refer to as *latent response classes*. The non-response model (6) is then given by

$$p(\mathbf{R}|\eta, \mathbf{X}) = \sum_{c=1}^{C} \left\{ \prod_{j=1}^{p} p(R_j|\xi = c, \eta, \mathbf{X}) \right\} p(\xi = c|\eta, \mathbf{X}). \tag{7}$$

We specify the model for ξ given η and $\mathbf{X}$ as the multinomial logistic model

$$p(\xi = c|\eta, \mathbf{X}) = \frac{\exp\left(\zeta_{0c}^{(g)} + \zeta_{1c}^{(g)}\eta + \sum_{k=1}^{K} \zeta_{2kc}^{(g)} X_k \right)}{\sum_{c'=1}^{C} \exp\left(\zeta_{0c'}^{(g)} + \zeta_{1c'}^{(g)}\eta + \sum_{k=1}^{K} \zeta_{2kc'}^{(g)} X_k \right)} \tag{8}$$

for $c = 1, \ldots, C$ in groups $g = 1, \ldots, G$ (this notation assumes that $\mathbf{X} = (X_1, \ldots, X_K)$ and that η is continuous). For identification, we set $\zeta_{01}^{(g)} = \zeta_{11}^{(g)} = \zeta_{2k1}^{(g)} = 0$ for all g and k. This could be extended to include also interactions between $\mathbf{X}$ and η and, say, quadratic effects in η. The models for R_j are binary regression models, which we specify as logit models of the form

$$\text{logit}\{p(R_j = 1|\xi, \eta, \mathbf{X})\} = \alpha_{0j}^{(g)} + \sum_{k=1}^{K} \alpha_{kj}^{(g)} X_k + \sum_{c=2}^{C} \beta_{0jc}^{(g)} \xi_{(c)} + \sum_{c=2}^{C} \sum_{k=1}^{K} \beta_{kjc}^{(g)} X_k \xi_{(c)} + \delta_{0j}^{(g)} \eta$$

$$+ \sum_{k=1}^{K} \delta_{kj}^{(g)} X_k \eta \tag{9}$$

for $j = 1, \ldots, p$, where the $\xi_{(c)}$ are dummy variables for the response classes $c = 2, \ldots, C$. If we impose the assumption $p(R_j|\xi, \eta, \mathbf{X}) = p(R_j|\xi, \mathbf{X})$ which was discussed in Section 3.1, $\delta_{0j}^{(g)} = \delta_{kj}^{(g)} = 0$ for all j, k and g. The hypothesis of ignorable non-response in group g is then simply that $\zeta_{1c}^{(g)} = 0$ for all c in model (8). The remaining parameters of model (9) describe the baseline log-odds that a respondent in group g responds to item j ($\alpha_{0j}^{(g)}$) and how these log-odds are affected by the covariates ($\alpha_{kj}^{(g)}$), the latent response classes ($\beta_{0jc}^{(g)}$) and interactions between them ($\beta_{kjc}^{(g)}$). For example, if $\beta_{0jc}^{(g)} + \Sigma_k \beta_{kjc}^{(g)} X_k$ is positive, individuals in response class c are more likely to respond to item j than are otherwise similar individuals in class 1. If we assume that $\alpha_{kj}^{(g)} = \beta_{kjc}^{(g)} = 0$ for all j, k, g and c, the non-response behaviour given the latent response class does not depend on the covariates $\mathbf{X}$. The joint model defined by equations (5) and (7)–(9) is identified if, first, the combined structural and measurement model for η is identified under ignorable non-response (which for model (2)–(3), for example, requires that $p \geqslant 3$)

and, second, if the latent class model (7)–(9) is identified when η is known (which requires that $C \leqslant 2^p/(1+p)$).

The latent class model (8)–(9) has potential advantages over models with continuous ξ. It avoids distributional assumptions such as assuming a normally distributed ξ. It also makes it easier to specify flexible and well fitting models for the joint distribution of **R**. If a single continuous ξ is not adequate, the fit of the model can be improved by using multiple ξ-variables (Rose, 2013; Köhler *et al.*, 2015a; Glas *et al.*, 2015; Rose *et al.*, 2015) or by allowing conditional dependences between some of the R_j. Both of these approaches, however, can be inconvenient and computationally demanding. With a categorical ξ, in contrast, the flexibility of the association structure can be increased simply by increasing the number C of the latent classes, if needed up to a point where an essentially saturated model for **R** is obtained. For example, individuals who have a high probability of responding for one kind of item and a low probability for another kind could be represented by one or more response classes with this characteristic, rather than with a high value for one continuous ξ and a low value for another.

We could also want to extend the model to allow for distinct response probabilities for different types of non-response, such as refusals and 'don't know' responses. This would mean defining R_j to have three or more unordered categories. The latent class model would handle this extension with no added difficulty, by changing it to include a multinomial model for unordered R_j. With continuous ξ, however, such a multinomial logistic model is less usual and would introduce additional computational complexity.

3.3. Estimation and model comparison

Let $\boldsymbol{\theta}$ denote the vector of all the parameters of the joint distribution (5). Given data on n units, which we assume to be independent, and adding now the unit subscript i to the notation, the log-likelihood function for $\boldsymbol{\theta}$ is

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\{p(\mathbf{R}_i, \mathbf{Y}_i|\mathbf{X}_i)\} \tag{10}$$

$$= \sum_{i=1}^{n} \log\left[\int \sum_{c=1}^{C} \left\{\prod_{j=1}^{p} p(R_{ji}|\xi_i=c,\eta_i,\mathbf{X}_i)\right\}\left\{\prod_{j\in\mathcal{O}_i} p(Y_{ji}|\eta_i)\right\} p(\xi_i=c|\eta_i,\mathbf{X}_i) p(\eta_i|\mathbf{X}_i) \,\mathrm{d}\eta_i\right]$$

where the structural and measurement models for Y_{ji} are specified as described in Section 2 and the latent response class non-response model as in Section 3.2 (if a different kind of response propensity model was used, the form of the likelihood would be similar, except that the corresponding form of model (6) would be substituted and the sum over ξ would be an integral if ξ was continuous). Maximizing $l(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ gives maximum likelihood estimates of the parameters. In this, the number of response classes C is taken as fixed. For selecting it, Akaike information criterion AIC- and Bayesian information criterion BIC-statistics, say, may be used to compare models with various choices for C. Likelihood ratio tests can be used for comparisons of nested models, in particular for testing the hypothesis of ignorable non-response (which is $p(\xi_i|\eta_i,\mathbf{X}_i) = p(\xi_i|\mathbf{X}_i)$ if we make the assumption that the R_{ji} do not depend directly on η_i).

Maximization of equation (10) requires an iterative algorithm, plus numerical methods of integration if ξ or η is continuous. We recommend the use of existing software packages which implement robust algorithms for general latent variable modelling. In our analysis we have used the Mplus software (Muthén and Muthén (2010) and Muthén (2004); examples of the code are given in the supplementary materials to this paper). It employs the EM algorithm, accelerated by occasional quasi-Newton and Fisher scoring optimization steps. Here each iteration of the EM

algorithm involves the evaluation and maximization of $\Sigma_i E[\log\{p(\mathbf{R}_i, \mathbf{Y}_i, \xi_i, \eta_i | \mathbf{X}_i)\}]$, where the expectation is with respect to $p(\xi_i, \eta_i | \mathbf{R}_i, \mathbf{Y}_i, \mathbf{X}_i)$ given the parameter values from the preceding iteration. Because log-likelihoods for latent class models commonly have multiple local maxima, the estimation was carried from multiple starting values, and with parameter estimates from simpler models used to generate starting values for more complex models.

4. Previous literature on latent response propensity models

The idea of using latent variables to model non-response has appeared in various guises in several literatures. In this section we provide a summary of these previously proposed methods, to relate them to each other and to our models defined in Section 3. For the still larger literature on models for non-ignorable non-response which do not invoke latent variables, see Molenberghs and Kenward (2007), part IV, and Little and Rubin (2002), chapter 15, and references therein.

A latent response propensity ξ (in our notation, which is used throughout this section) appears already in sample selection models for single items Y_j which are not treated as measures of latent variables η, going back to at least Heckman (1976) and Nelson (1977). The earliest applications of this idea to multivariate responses were to longitudinal data. There the variables $\mathbf{Y}_C$ are the intended repeated measurements of a response variable, and η are the random effects in the model for $\mathbf{Y}_C$. The kind of non-response that has been considered most often is dropout, where the elements of $\mathbf{Y}_C$ for a respondent are all observed up to some time and all missing after it. In this case, the response indicators $\mathbf{R}$ can be replaced with a single variable R for the time of dropout, and a survival analysis model defined for R. Models where both $\mathbf{Y}_C$ and R depend on η are known as *shared parameter models* for dropout. They were first proposed by Wu and Carroll (1988) and further developed by, among others, Schluchter (1992), De Gruttola and Tu (1994) and Ten Have *et al.* (1998), and, for pattern–mixture extensions where the distribution of $\mathbf{Y}_C$ depends also on R, by Wu and Bailey (1989), Hogan and Laird (1997) and Hogan *et al.* (2004). Further references and overviews of this by now large literature can be found in Little (1995), Daniels and Hogan (2008) and Ibrahim and Molenberghs (2009).

Although it has focused on models where the non-response is dropout, η is continuous and the same η is used to predict both $\mathbf{Y}_C$ and R, the literature on longitudinal data has also examined other situations which are closer to the case that we consider for multivariate item data. The possibility of intermittent patterns of non-response was allowed already by Follmann and Wu (1995), and by Cowles *et al.* (1996) who also introduced additional continuous latent variables ξ to model the non-response. Models with a categorical ξ for longitudinal data were proposed by Roy (2003), Lin *et al.* (2004) and Beunckens *et al.* (2008) (see also Muthén *et al.* (2003)). Hafez *et al.* (2015) considered models for dropout where the outcomes that are observed at each occasion are themselves sets of multiple items, and Glas and Pimentel (2008) models for dropout-like monotone patterns of non-response in multiple items in cross-sectional data.

For multiple-item data, the idea of latent response propensities appears implicitly in Lord (1974, 1983) and explicitly in Muthén *et al.* (1987) who stated how such models could be defined in the context of factor analysis. Knott *et al.* (1990) were the first to implement and estimate models like this, in their case for binary items $\mathbf{Y}_C$; their non-response models are of the form $p(R_j | \xi, \eta)$ where ξ and η are univariate and independent of each other. O'Muircheartaigh and Moustaki (1999) generalized this to situations where $\mathbf{Y}_C$ may include both continuous and binary items, and Moustaki and Knott (2000) to the case where the non-response and measurement models (but not the structural model) may also depend on covariates $\mathbf{X}$. Holman and Glas (2005) introduced models where $p(R_j | \xi)$ depend only on the response propensities ξ, but η and ξ may be correlated, and Rose *et al.* (2015) and Glas *et al.* (2015) extended these to allow for

covariates in the models for these latent variables. In all of these models, η and ξ are continuous variables.

In recent years such models have been studied most actively in the context of item response theory modelling of data from educational and psychological testing, where the items are categorical (often binary) responses to test questions, the substantive latent variables η describe respondents' abilities or psychological characteristics and non-ignorable non-response may arise when respondents omit questions by choice or because of running out of time, for reasons which may be related to η (Mislevy and Wu, 1996; Mislevy, 2016). Developments, applications and evaluations of latent response propensity models in this field include Glas and Pimentel (2008), Korobko *et al.* (2008), Bertoli-Barsotti and Punzo (2013), Pohl *et al.* (2014) and Köhler *et al.* (2015a,b). Of particular relevance for the cross-national focus of our paper is Rose *et al.* (2010), who described different approaches to modelling cross-national data from educational tests with non-response, and carried out a multigroup analysis with latent response propensities for data from the Programme for International Student Assesssment in 30 countries. Their modelling uses only non-response models $p(R_j|\xi)$ which are equivalent across the countries.

Models for multiple items where the response propensity ξ is a categorical variable, which are our focus, have received less attention. Bacci and Bartolucci (2015) defined such models where η is also categorical, η and ξ are conditionally independent given $\mathbf{X}$ and the non-response model may depend on both η and ξ. Jung *et al.* (2011) also used a categorical ξ but combined with a model for the joint distribution of $\mathbf{Y}_C$ which is not itself a latent variable model.

Two further approaches are closely related to models with latent response propensities ξ. One of them is to use the same latent η to model both $\mathbf{Y}_C$ and $\mathbf{R}$, without introducing a separate ξ. This can be done by treating $\mathbf{R}$ as additional observed items alongside $\mathbf{Y}_C$ (Harel and Schafer, 2009) or by including 'missing' as an additional category for each (categorical) item Y_j (Bock, 1972; Moustaki and O'Muircheartaigh, 2000; Formann, 2007). The second alternative approach omits ξ but uses some function $f(\mathbf{R})$ of $\mathbf{R}$—e.g. the observed response rate $\bar{R} = p^{-1}\Sigma_j R_j$ for a respondent—as a proxy measure for it. Models like this were proposed by Rose *et al.* (2010, 2015) and (in a different context) by Yuan and Little (2007). This approach is fully valid if $\mathbf{Y}_{mis}$ and $\mathbf{R}$ are conditionally independent given $(\mathbf{Y}, \mathbf{X}, f(\mathbf{R}))$, which is a situation which Harel and Schafer (2009) called 'partially missing at random'.

Further variants of these ideas have also been applied in survey sampling analysis where the goal is to estimate population characteristics of a single variable Y_j. Models with latent response propensities can be brought to bear even on this task if we can define the vector $\mathbf{Y}_C$ in a helpful way. Moustaki and Knott (2000) and Matei and Ranalli (2015) did this by taking $\mathbf{Y}_C$ to be a set of multiple variables, including but not limited to Y_j, and fitting a model for their response indicators $\mathbf{R}$ given ξ and (optionally) $\mathbf{X}$. They then derived from this model an estimated response probability of $R_j = 1$ for Y_j for each respondent in the observed sample and used these to obtain non-response weights for estimating population quantities of Y_j. Another context to which the same ideas can be adapted is cluster sampling, where a cluster level random effect is in the role of η and the vector of the Y_j for all the sample members in a cluster acts as $\mathbf{Y}_C$ (Shao, 2007; Yuan and Little, 2007, 2008; Skinner and D'Arrigo, 2011).

5. Analysis of welfare attitudes in the European Social Survey

5.1. Data and variables

We analyse data from 29 countries in round 4 of the ESS (European Social Survey, 2008). As the items $\mathbf{Y}_C$, we consider three questions from the ESS module on 'Welfare attitudes in a changing Europe' (Svalfors *et al.*, 2008). The items, the labels which we use to refer to them and their

Table 1. The three attitude items (with wordings in English, and labels which will be used to refer to the items) and individual level covariates in the data from round 4 of the ESS which are used in the analyses of Section 5

Attitude items, measuring attitudes towards receivers of welfare benefits
'[...] Please say how much you agree or disagree with each of the following statements about people in [country]'

D40	'Most unemployed people do not really try to find a job'
D42	'Many people manage to obtain benefits and services to which they are not entitled'
D44	'Employees often pretend they are sick in order to stay at home'
Response options	Agree strongly (1); agree (2); neither agree nor disagree (3); disagree (4); disagree strongly (5)

Respondent level covariates

Sex	Dummy variable for women
Age (and Age^2)	In 10s of years, centred on 50 years
Education	Highest level of education obtained, in three categories: lower secondary or less; upper secondary; post secondary†

---

†See European Social Survey Data Archive (2009) for information on the cross-nationally comparable coding of education.

response options are shown in Table 1. The items are regarded as measures of a continuous latent variable η, which will be interpreted as the perceived level of abuse of the welfare system by its users. We treat the items as interval level variables and use the factor analysis model (2)–(3) for η. With three items, the measurement model is saturated, so its goodness of fit is not a concern. The analysis also includes individual level covariates **X** on the respondent's age, age squared, sex and education, coded as shown in Table 1.

Table 2 shows the countries, sample sizes and non-response rates in the data. The analysis uses all respondents with data on all three covariates, omitting the very small number with missing values in any of them. For the attitude items, responses coded as 'don't know', 'refusal' or 'no answer' are treated as item non-response. Rates of non-response vary between 0.2% for item D40 in Norway and Belgium, and 20.2% for D44 in Bulgaria. In all the countries combined, 10.5% of the respondents had a missing value for at least one of the three items, and this figure varied from 1.0% in Norway to 27.3% in Bulgaria.

We carry out analyses which demonstrate the use and interpretation of the models that were introduced in Sections 2 and 3, with the non-response model specified as the latent response class model (7)–(9). In Section 5.2 we first estimate models for each of the 29 countries separately. This allows qualitative comparisons of the results between the countries, without any cross-national constraints on the model parameters. In Section 5.3 we then move to multigroup models which are fitted jointly for several countries. We describe two such examples: one with a model with covariates for a subset of the countries, and one without covariates for all 29 countries. Command files for replicating the analyses in this section are available from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

5.2. *Country-by-country analyses*

When the models are fitted separately for each country, all of the parameters of all parts of the model vary freely between countries. For identifiability of the scale of η, we set $\gamma_0^{(g)} = 0$ and $\psi^{(g)} = 1$ in each country-specific structural model (2). We make throughout the additional assumption that $p(R_j|\xi, \mathbf{X}) = p(R_j|\xi)$, i.e. that the response indicators R_j do not depend directly on any other covariates than the country.

Table 2. List of the countries in round 4 of the ESS†

Country	Non-response rate (%)				Sample size
	Item D40	Item D42	Item D44	At least one item	
Finland (FIN)	0.4	1.2	0.8	1.7	2194
Sweden (SWE)	0.7	4.5	2.3	5.9	1822
Norway (NOR)	0.2	0.9	0.3	1.0	1543
Denmark (DEN)	0.9	5.2	2.3	6.6	1600
Ireland (IRL)	0.5	1.7	0.9	2.5	1749
United Kingdom (UK)	1.1	1.9	1.5	3.4	2321
Belgium (BEL)	0.2	1.9	0.6	2.4	1751
Netherlands (NET)	2.1	4.9	2.9	7.3	1777
France (FRA)	0.4	1.6	0.9	2.5	2071
Germany (GER)	1.4	3.4	2.0	5.4	2725
Switzerland (SWI)	1.3	5.2	2.8	8.1	1815
Portugal (POR)	3.7	8.1	8.9	14.1	2366
Spain (SPA)	2.5	8.4	5.0	12.2	2570
Greece (GRE)	1.6	1.8	3.7	5.3	2070
Cyprus (CYP)	4.1	9.0	12.5	17.6	1213
Turkey (TUR)	3.5	3.4	8.2	9.6	2386
Israel (ISR)	3.8	10.7	9.2	15.7	2442
Estonia (EST)	2.6	15.6	8.0	19.6	1656
Latvia (LAT)	1.3	7.3	5.6	10.6	1980
Poland (POL)	1.7	5.8	6.1	10.2	1616
Czech Republic (CZE)	1.4	8.3	5.5	11.4	2015
Slovakia (SLK)	1.8	6.8	6.8	11.4	1779
Slovenia (SLN)	2.1	5.2	5.9	9.6	1283
Croatia (CRO)	2.8	3.3	6.2	8.7	1443
Hungary (HUN)	1.0	4.2	5.9	8.6	1543
Bulgaria (BUL)	4.8	13.5	20.2	27.3	2230
Romania (ROM)	5.3	9.3	12.6	17.1	2093
Ukraine (UKR)	3.8	12.2	14.8	22.9	1840
Russia (RUS)	3.5	11.5	13.6	21.8	2508
All countries	2.1	6.2	6.2	10.5	56401

†The countries are listed in a rough geographical order. 'Sample size' denotes the number of observations that were used in the analyses, i.e the number of respondents for whom all the covariates (sex, age and education) are observed. Also shown are percentages, out of these samples, of missing values for each of the three attitude items (as listed in Table 1) and the percentage of respondents for whom at least one of these items is missing.

In general, we suggest that the first step should be to choose the number of the latent response classes (C), setting it sufficiently high that it explains well the dependences among $\mathbf{R}$. Models with different values of C can be compared by using, for example, the AIC- or BIC-statistics. To simplify the computing, it may often be convenient to carry out the selection of the number of classes by fitting standard latent class models first for $\mathbf{R}$ only, ignoring $\mathbf{Y}_C$ and the covariates $\mathbf{X}$. In our application this step is redundant because, with three items, a two-class model already produces a saturated model for the joint distribution of $\mathbf{R} = (R_1, R_2, R_3)$. The interpretation of the two latent classes is similar for all countries. Class 1 is the class of respondents, where the probability of answering each of the items is nearly 1, whereas class 2 can be interpreted as the potential non-respondents. For example, for the UK the estimated probability that a respondent answers all three items is 0.996 in class 1 and 0.332 in class 2.

Table 3. Parameter estimates for the measurement and structural models for positive attitude towards recipients of welfare benefits, fitted separately for each country†

Country‡	Factor loading λ_j for the following items:			Coefficients of covariates in the structural model for latent attitude variable η				
	D40	D42	D44	Age	Age^2	Woman	Education	
							Upper secondary	Post secondary
FIN	0.62	0.50	0.61	−0.08*	−0.03*	−0.05	0.18§§	0.65*
SWE	0.50	0.38	0.63	0.06*	−0.06*	0.14§	0.22§§	0.74*
NOR	0.46	0.45	0.60	0.02	−0.04*	0.14§	0.07	0.59*
DEN	0.54	0.45	0.59	−0.02	−0.03§§	0.08	0.48*	0.93*
IRL	0.45	0.41	0.48	−0.01	−0.03§§	−0.09	0.23§	0.50*
UK	0.53	0.48	0.48	0.00	−0.00	−0.04	0.11	0.45*
BEL	0.63	0.48	0.52	−0.03	−0.01	−0.05	0.11	0.62*
NET	0.59	0.50	0.55	0.04§	−0.03§§	0.06	0.34*	0.84*
FRA	0.80	0.59	0.71	−0.03§	−0.03§§	−0.05	0.24*	0.84*
GER	0.67	0.39	0.57	0.02	−0.03*	0.06	0.18§	0.59*
SWI	0.67	0.56	0.64	−0.05§§	−0.03§§	−0.09	0.15§	0.65*
POR	0.75	0.54	0.63	−0.00	−0.00	0.02	0.07	0.28§§
SPA	0.43	0.37	0.81	0.01	−0.00	0.05	0.16§	0.18§
GRE	0.37	0.20	1.10	0.02	−0.01	0.01	0.09	0.16§
CYP	0.64	0.42	0.43	−0.01	0.01	−0.00	−0.01	0.10
TUR	0.57	0.26	0.87	0.01	0.00	0.02	−0.02	−0.08
ISR	0.38	0.45	0.86	0.01	0.01	−0.04	−0.29*	−0.15§
EST	0.62	0.32	0.59	0.01	0.02	0.09	0.19§	0.52*
LAT	0.71	0.46	0.58	−0.01	−0.03§§	0.05	−0.17§	−0.23§
POL	0.53	0.39	0.49	0.02	0.01	0.16§	−0.07	−0.07
CZE	0.73	0.49	0.63	−0.01	−0.01	−0.05	−0.04	−0.11
SLK	0.66	0.58	0.36	−0.05§§	−0.00	0.02	−0.06	−0.18
SLN	0.63	0.52	0.54	−0.03	0.01	−0.04	0.01	0.45*
CRO	0.53	0.23	0.84	0.02	−0.02	0.09	−0.01	−0.02
HUN	0.66	0.37	0.70	−0.03	0.00	0.15§	0.07	−0.05
BUL	0.74	0.37	0.70	−0.02	0.01	0.05	−0.39*	−0.35*
ROM	0.61	0.34	0.87	0.03	−0.02	0.03	−0.07	−0.14
UKR	0.65	0.47	1.05	0.06*	0.01	0.03	0.26§§	0.16
RUS	0.47	0.41	0.78	0.01	−0.01	0.14§	−0.07	0.06

†The estimates are from joint models which also include a non-response model which allows for non-ignorable non-response (see Table 4). The variables are as defined in Table 1. All factor loadings have at least $p < 0.05$.
‡See Table 2 for explanation of the country abbreviations.
§$p < 0.05$.
§§$p < 0.01$.
*$p < 0.001$.

The parameter estimates for each country are given in Tables 3 and 4. Table 3 shows the factor loadings of the measurement models of the items given the latent attitude η (λ_j in equation (3)) and the regression coefficients of the covariates in the structural model for η ($\gamma_x^{(g)}$ in (2)). For every country, larger values of η indicate more positive attitudes towards the receivers of welfare provision. The value 0 for η corresponds to the average attitude in a country for men aged 50 years who have at most lower secondary education. In the structural models, sex has a significant association with the attitude (at the 5% level of significance) in five countries, age in under half of the countries and education in more than half of them. Women in Sweden,

Norway, Poland, Hungary and Russia have on average a more positive attitude towards welfare recipients than do men. In most countries higher levels of education are associated with more positive attitudes, with Bulgaria and Israel the only countries where this association is significant in the opposite direction. For age, the model includes a quadratic term. Its significant negative coefficients show that middle-aged respondents tend to have more positive attitudes than the young and the old, with the highest expected values typically around ages 45–55 years. There is also a clear geographic pattern in that most of the significant associations with age and education are observed in countries in northern and western Europe.

Table 4 shows the estimated coefficients of model (8) for the non-response class ξ given attitude η and the covariates. The coefficients are log-odds ratios of being in the class of item non-respondents rather than the class of respondents, and the intercept term gives the log-odds of being in the non-response class for a man aged 50 years who has at most lower secondary education and value 0 for the attitude. The conclusions about significant predictors of non-response are consistent in direction, i.e. for every covariate the coefficient has the same sign in every country where it is significant; these results are in fact more consistent than those for the attitude variable, where there were significant associations in both directions for age and education. Women and people with the least education are more likely to be non-respondents. The quadratic relationship between age and response propensity is such that the middle-aged are more likely to be respondents than are the young and the old. This relationship is significant in all except five of the countries, and very similar across many of them: for example, in 16 countries the highest probability of the respondent class is estimated to occur between 45 and 50 years of age. Comparing the results for age between Tables 3 and 4, we observe that for many countries we might thus include the quadratic age effect in the model for the response class ξ but omit it from the structural model for η; this illustrates the general point that the functional forms in which covariates **X** enter these models need not be the same.

For the main focus of this paper, the most interesting predictor in the non-response model is the latent attitude η. In 19 of the 29 countries it has a significant association with the latent response class, meaning that non-response is non-ignorable even after controlling for age, sex and education of the respondents. The strength of this association is in many countries such that a difference of 1–2 (residual) standard deviations in η has an approximately comparable association with the non-response probability than does the difference between the highest and lowest levels of education. The significant associations are again consistent in direction across these 19 countries: in all of them, individuals with more positive attitudes towards welfare recipients are less likely to answer these survey questions. There is thus no evidence of a social desirability bias in the direction of not reporting negative attitudes, but evidence of the opposite tendency. There is also some geographic pattern in these results, in that the non-response is most consistently and strongly non-ignorable in the countries of central and eastern Europe.

The model for ξ could also include interactions or non-linear terms involving the latent η. Here very few such terms were significant so they are not included in the results in Tables 3 and 4, but we mention this possibility here to illustrate what such effects would mean. For example, a quadratic effect of η would capture a pattern where the probability of non-response was higher (or lower) for those with extreme values of an attitude η—whether extremely positive or extremely negative—than for those with non-extreme levels of it. Interactions between η and covariates could indicate that the strength of non-ignorability of non-response depended on respondent characteristics, e.g. if the probability of non-response depended less strongly on the attitude for respondents with higher levels of education.

Table 4. Parameter estimates for the model for the latent response class ξ given covariates and the latent η (positive attitude towards welfare recipients), fitted separately for each country†

Country‡	Intercept	Age	Age^2	Woman	Education		Attitude η
					Upper secondary	Post secondary	
FIN	−4.67	0.41§	0.04	0.73	−1.08	−0.95	−0.07
SWE	−4.05	0.02	0.16§§	0.34	−0.99	−1.22	0.25
NOR	−5.65	0.19	0.21§	0.19	−1.31	(−24.8)	−0.38
DEN	−5.28	0.15	0.24§§	0.85	−1.09	−0.88§	0.39
IRL	−3.29	0.06	0.10§	−0.15	0.01	−0.85	0.60§
UK	−3.53	0.18§	0.08§	0.45	−2.17§	−0.25	0.45§
BEL	−3.59	0.09	0.07	0.55	0.25	−0.26	0.72
NET	−4.12	−0.08	0.12§	0.03	−1.51§§	−1.55§	1.13§
FRA	−3.64	0.40§§	−0.03	0.55	−0.17	−0.67	0.53§
GER	−2.95	0.10	0.12§§	0.32	−1.06§§	−1.45§§	0.52§
SWI	−3.55	0.15	0.16§§	0.97§	−0.87	−0.61	0.68§
POR	−3.36	0.09	0.17*	0.36	−0.33	−0.55	0.54§§
SPA	−3.00	0.15§§	0.15*	0.57§§	−0.78§	−1.37§§	0.71*
GRE	−3.49	0.02	0.10§	0.84§	−1.51§§	−2.00§§	0.02
CYP	−2.65	0.30*	0.08§	0.83§§	−0.56	−0.79§	−0.01
TUR	−3.02	0.14§	0.07§§	0.56§§	−0.03	−0.86	0.32
ISR	−2.21	0.06	0.07§§	0.22	−0.95*	−0.72§§	1.62*
EST	−1.97	0.35*	0.03	0.37	−0.22	−0.70§§	0.35§
LAT	−3.73	0.14§	0.08§	0.37	0.09	−0.33	0.30
POL	−2.43	0.24*	0.09§§	0.56§	−1.25*	−0.99§§	0.76§§
CZE	−3.48	0.10	0.12§§	0.24	−0.24	−0.09	0.41
SLK	−4.10	0.06	0.14§§	0.45	−0.31	0.05	0.73§§
SLN	−2.85	0.36*	0.07	0.05	−0.60	−1.12§	0.71§§
CRO	−3.51	0.14	0.11§	0.08	−0.54	0.08	1.17§§
HUN	−4.48	0.14	0.19*	−0.58	−0.23	−1.34	1.34§§
BUL	−1.22	−0.01	0.12*	0.02	−0.79*	−0.93*	0.88*
ROM	−2.53	0.08	0.12*	0.41§	−0.86*	−0.93§§	0.53§
UKR	−2.26	0.02	0.08§§	0.50§	−0.39	−0.67§	0.64§§
RUS	−2.40	0.05	0.07§	0.37	−1.04§§	−1.04*	0.57§

†Each model is a binary logistic model for being in the class of non-respondents, so a positive regression coefficient indicates increased probability of being a non-respondent.
‡See Table 2 for explanation of the country abbreviations.
§$p < 0.05$.
§§$p < 0.01$.
*$p < 0.001$.

5.3. Multigroup analyses

The first multigroup analysis considers six countries: the UK, Germany, Sweden, Spain, Poland and Bulgaria. They have different levels of non-response (see Table 2) and strengths of non-ignorability in it (Table 4). Substantively, this selection of countries is motivated by the typology of welfare state regimes by Esping-Andersen (1990) (we draw also on Ferragina and Seeleib-Kaiser (2011)). The six countries represent different regime types, and individuals in them might for that reason display different attitudes towards welfare recipients: the UK, Germany and Sweden are clear instances of liberal, Christian democratic and social democratic welfare states respectively, whereas Spain (from the Mediterranean) and Poland (from central–eastern Europe) represent regions of Europe whose regimes are not fully captured by the original typology.

Bulgaria (a second central–eastern European country) was added because it has the highest level of non-response for these items in the ESS sample.

The types of models that are fitted here are the same as in Section 5.2. The non-response model is again a two-class latent class model conditional on η and the covariates, and all of its parameters are allowed to vary across the countries. The response classes in each country can again be interpreted as those of respondents and potential item non-respondents. Estimated coefficients of the model for the class membership are shown in the uppermost part of Table 5. The conclusions from them are very similar to those from the country-specific analyses in Table 4.

Cross-national measurement equivalence is assumed in the measurement model of the items. In the structural model, we take the UK as the reference country with a fixed intercept term $\gamma_0^{(1)} = 0$. The residual variance $\psi^{(g)}$ is fixed at 1 for all countries for simplicity (this also makes the scale of the latent η similar in this respect to those in the country-specific analyses). All other parameters of the structural model may vary freely across the six countries. Parameter estimates for this model are shown in the middle part of Table 5. Substantive interest in cross-national survey analysis would typically centre on this model. We might first examine interactions between country and the covariates, in other words how the effects of the covariates on the attitude may vary between the countries. Here this interaction was not significant between country and gender ($p = 0.48$ for a likelihood ratio test), so the gender coefficient is constrained to be equal across countries. Controlling for the other covariates, women tend to have slightly more positive attitudes towards welfare recipients than do men. The other covariates have significantly different coefficients in different countries. The results for them are similar to those from the country-specific analyses in Table 3, i.e. that age is either not associated or non-linearly associated with the attitude, and that in five of the six countries individuals with more education tend to have more positive attitudes.

A set of parameters in the structural model which can only be estimated from a joint multi-country analysis is the intercept terms $\gamma_0^{(g)}$. These describe expected differences in attitudes between individuals from different countries, on top of those explained by differences in the individuals' age, gender and education. Here there is substantial cross-national variation. After controlling for the covariates, expected attitudes towards welfare recipients are most negative in the UK and Poland, and most positive in Bulgaria and Sweden. These differences are substantially large (in several cases more than the within-country standard deviation of 1) and statistically significant for most pairs of countries.

Our second multicountry analysis includes all 29 countries in the ESS. Here the model that was considered above, which includes full interactions between the countries and all the covariates in both the structural and the non-response models, has a very large number of parameters, and it became computationally infeasible with our implementation of the estimation. Instead, for this example of joint analysis we examine a simpler model with country as its only covariate. This means that the only estimated parameters of the structural model are the country-specific intercepts $\gamma_0^{(g)}$. They are now interpreted as the average levels of the attitude towards welfare recipients within each country, rather than the averages after controlling for differences in the distributions of the covariates. In the non-response model we further assume that for each item the response probabilities given a latent class are the same across countries. This restriction, which implies some compromise of the goal of maximum flexibility of the non-response model, is also imposed for computational feasibility. The parameters of the model for the response class ξ given country and attitude η, in contrast, are still different for each country, to allow for different levels of non-response and non-ignorability across the countries.

Table 6 shows estimates of the key parameters from this analysis. The column labelled '$\eta \rightarrow \xi$' shows the coefficients of the attitude in the model for the response class. The cross-national

Table 5. Parameter estimates for models for positive attitude towards welfare recipients, fitted jointly to data from the UK, Germany, Sweden, Spain, Poland, and Bulgaria†

Country	Intercept (standard error)	Coefficients of explanatory variables					
		Age	Age^2	Woman	Education		Attitude η
					Upper secondary	Post secondary	
Model which assumes non-ignorable non-response							
Non-response model for the log-odds of being in the latent non-response class ($\xi = 2$)							
UK	−4.01	0.17§	0.10§§	0.49	−2.12§	−0.44	0.78§§
GER	−3.66	0.09	0.12§§	0.29	−1.06§§	−1.44§§	0.62§
SWE	−4.79	0.01	0.16§§	0.35	−1.03	−1.33	0.42
SPA	−3.57	0.17*	0.15*	0.57§§	−0.84§	−1.47*	0.81*
POL	−3.28	0.27*	0.12§§	0.74§	−1.44*	−1.41§§	0.82§§
BUL	−2.23	−0.01	0.11*	0.03	−0.75*	−0.91*	0.76*
Structural model for the expected value of the attitude η							
UK	0	0.01	−0.01	0.06§§	0.12	0.43*	
GER	1.13 (0.10)	0.03	−0.03*	0.06§§	0.16	0.56*	
SWE	1.64 (0.10)	0.06§§	−0.06*	0.06§§	0.20§§	0.72*	
SPA	0.65 (0.07)	−0.01	−0.00	0.06§§	0.21§§	0.25*	
POL	0.37 (0.11)	0.03	−0.01	0.06§§	−0.09	−0.06	
BUL	1.35 (0.09)	−0.02	0.00	0.06§§	−0.43*	−0.35*	
Model which assumes ignorable non-response							
Structural model for the expected value of the attitude η							
UK	0	0.01	−0.01	0.06§	0.13	0.44*	
GER	1.13 (0.11)	0.03	−0.04*	0.06§	0.17§	0.57*	
SWE	1.66 (0.10)	0.06§§	−0.06*	0.06§	0.21§§	0.73*	
SPA	0.64 (0.07)	−0.01	−0.00	0.06§	0.22§§	0.26*	
POL	0.36 (0.11)	0.03	−0.01	0.06§	−0.07	−0.04	
BUL	1.32 (0.09)	−0.02	−0.00	0.06§	−0.41*	−0.33*	

†The upper part of the table shows a model which also includes a latent class model for non-ignorable non-response in measures of the attitude, and the lower part a model which omits the non-response model (i.e. assumes ignorable non-response). In both structural models, the coefficient of Woman is constrained to be equal across countries. Measurement model for η in both models: estimated factor loadings for items D40, D42 and D44 are 0.52, 0.36 and 0.66 respectively, all significant to at least $p < 0.05$.
§$p < 0.05$.
§§$p < 0.01$.
*$p < 0.001$.

patterns in them are broadly comparable with those from the country-specific analyses with covariates. The coefficients are typically smaller in value here, and some of them lose statistical significance. It is thus not the case here that the non-response would be judged ignorable when the analysis is conditional on covariates and non-ignorable when it is not, and in fact the opposite tends to be so. This may be due to the particular patterns of associations between these variables. When controlling for the covariates, the latent attitude tends to be positively associated with the probability of non-response. Of the covariates, education is the strongest predictor, and it typically has a positive association with attitude but a negative association with non-response. This mixture of signs implies that, when we do not control for the covariates, the confounding by education tends to suppress the effect of attitude towards zero.

The column 'non-ignorable' in Table 6 shows the estimated mean attitudes by country ($\gamma_0^{(g)}$)

Table 6. Parameter estimates for models for positive attitude towards recipients of welfare benefits, fitted jointly for data on all the countries listed in Table 2†

Country‡	Mean attitude when non-response is modelled as follows:		$\eta \to \xi$	Country	Mean attitude when non-response is modelled as follows:		$\eta \to \xi$
	Ignorable	Non-ignorable			Ignorable	Non-ignorable	
FIN	1.03	1.01	−0.65	TUR	−0.10	−0.09	0.25
SWE	1.40	1.39	0.27	ISR	0.42	0.47	1.73*
NOR	1.09	1.07	(−12)	EST	1.06	1.06	0.26§
DEN	1.46	1.44	−0.29	LAT	1.05	1.04	0.05
IRL	0.21	0.21	1.11	POL	0.06	0.06	0.43
UK	0	0	0.72	CZE	0.54	0.53	0.14
BEL	0.51	0.50	0.73	SLK	0.51	0.50	−0.15
NET	0.90	0.89	0.31	SLN	0.21	0.22	0.71§§
FRA	0.74	0.73	0.40	CRO	0.22	0.24	1.10§§
GER	1.01	1.00	−0.10	HUN	−0.08	−0.07	0.64§§
SWI	1.03	1.02	0.30	BUL	0.75	0.77	0.41*
POR	0.51	0.51	0.38§§	ROM	0.27	0.27	0.23§
SPA	0.50	0.50	0.55*	UKR	0.91	0.91	0.20§
GRE	0.52	0.51	−0.34	RUS	0.94	0.92	−0.01
CYP	0.14	0.13	−0.14				

†For each country, the first two columns show the mean level of the attitude (on a scale where the mean for the UK is 0), estimated from a model which assumes ignorable (first column) and non-ignorable (second column) non-response. The column '$\eta \to \xi$' shows the estimated coefficient of attitude on response class in the non-ignorable non-response model. Significance levels of the mean attitudes are not shown. Their standard errors are around 0.04 from both the ignorable and the non-inorable models. Measurement model for η: estimated factor loadings for items D40, D42 and D44 are 0.54, 0.47 and 0.66 respectively in the non-ignorable model, all significant to at least $p < 0.05$.
‡See Table 2 for explanation of country abbreviations.
§$p < 0.05$.
§§$p < 0.01$.
*$p < 0.001$.

from this model where the non-response model is taken to be non-ignorable (the estimates in the 'ignorable' column are discussed in the next section). These estimates are also shown in graphical form on the horizontal axis of Fig. 2. They can be used to make cross-national comparisons of average attitudes towards welfare recipients, as measured by these survey items. The range of these country means is substantively large, with the difference between the countries with the most positive attitudes on average (Denmark and Sweden) and those with the most negative attitudes (Turkey, Hungary and the UK) being around 1.5 individual level standard deviations in the attitude. The standard errors of the estimated means are around 0.04, so most of the pairwise differences between the countries are statistically significant.

5.4. Effects of allowing for non-ignorable non-response
The results that were reported above include examples of how response propensity models can be used to examine levels and patterns of non-response. Typically, however, a more important goal of the analysis is to estimate measurement and structural models for the latent attitude variables, and the main role of the non-response model is just to allow for the non-response in such a way that valid conclusions can be drawn about those models of main interest. The question that then arises is how sensitive these conclusions are to different specifications of
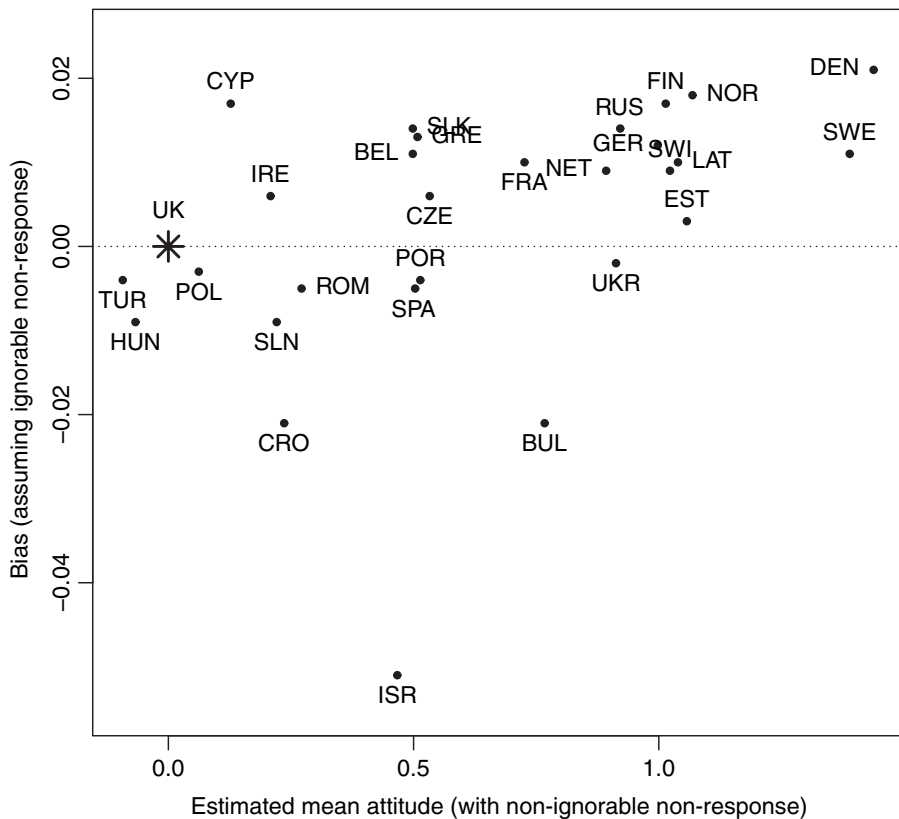
Fig. 2. Estimated mean attitudes by country from the model shown in Table 6 under various assumptions about the non-reponse (with the mean for the UK fixed at 0, and larger values indicating more positive attitudes): here the horizontal axis shows the estimates when the non-response is assumed non-ignorable, and the vertical axis the difference between the estimates under ignorable *versus* non-ignorable non-response

the non-response model. In particular, will the estimates be very different if we assume that the non-response is ignorable (i.e. we omit the non-response model), even when it is clearly non-ignorable?

Sensitivity analyses of this question have been reported in recent literature on response propensity models in the context of educational and psychological testing, using both empirical examples and simulation studies with settings motivated by such applications (Holman and Glas, 2005; Rose, 2013; Bertoli-Barsotti and Punzo, 2013; Pohl *et al.*, 2014; Köhler *et al.*, 2015a; Glas *et al.*, 2015). A common conclusion from these studies is that estimated parameters of the measurement models of the items are not seriously affected by incorrectly assuming that non-response is ignorable, unless the non-ignorability is strong and the rate of non-response is as high as 30–50%. Predicted values ('factor scores') for the latent variables η are also unaffected on average, but individual values of them (specifically those for respondents with extreme values of η) can be fairly sensitive to different assumptions about the non-response model, and the predictions can be more precise when information from the non-response indicators is included.

These previous studies have not considered the sensitivity of estimated structural models, which are the main focus of our analyses. An important exception is the empirical study of Rose

et al. (2010), on the Programme for International Student Assesssment. Their example is very relevant for our paper as well because it also uses cross-national data, with a similar number of countries to that in our ESS example. The educational test items in the programme's data and the survey items in our example have roughly comparable non-response rates, and the non-response is clearly non-ignorable in both cases. For a multigroup model without covariates, Rose *et al.* (2010) concluded that the estimated means of latent (ability) variables η are very similar when a non-ignorable non-response model is included and when it is not.

Conclusions from our application are similar. For the first multigroup example of Section 5.3, the bottom part of Table 5 shows estimates for the structural model estimated under the assumption of ignorable non-response. Here the effect is small, in that the parameter estimates, their standard errors (which are not shown in Table 5) and conclusions about country-by-covariate interactions are all effectively unchanged compared with those from the joint model under non-ignorable non-response. The differences in the two sets of estimated intercepts are in an expected direction: countries with the highest rates of non-response have larger intercepts (i.e. are more different from the UK) when we allow for non-ignorable non-response, as we would expect given that the estimated non-ignorability is such that a missing response indicates a more positive attitude. However, these differences are small.

For our second multigroup example, the column 'ignorable' in Table 6 lists the estimated mean attitudes by country when the non-response model is taken to be ignorable, and the vertical axis of Fig. 2 shows the differences between these and the means estimated under non-ignorable non-response. This difference is positively correlated with the mean attitude, so the overall effect of assuming ignorable non-response would be to exaggerate the differences between the countries. The largest individual biases are for Israel, Croatia and Bulgaria. These are countries where the general non-response rate is high and/or the non-response is strongly indicative of more positive attitudes, so assuming ignorability leads to an underestimate of the mean attitude. Even for these countries, however, the differences are again small. The estimated standard errors of the country means are slightly smaller in the non-ignorable non-response model.

In the kind of multivariate analysis that we consider here, the models use both the observed items **Y** and the response indicators **R** to provide information on individuals' values of the latent attitude η. For example, in our application the direction of non-ignorability was such that item non-response would indicate a more positive level of the attitude towards welfare recipients than would be concluded from an individual's recorded answers to the items alone. The relative contributions of these data are typically not equal, in that we would expect the items to provide most of the information and the indicators a smaller amount of it. In this situation, incorrectly ignoring **R** need not have a substantial influence on estimates of models for η. It would be most likely to have such an effect when **R** provides a relatively high proportion of the information on η, i.e. when non-response is very common (so that for many respondents very few or none of the items are observed) or very strongly non-ignorable. In our example the non-ignorability was moderately strong, but the non-response rate was fairly low.

These results suggest that in many contexts a latent variable model which assumes only ignorable item non-response may be quite adequate, even when the non-response is truly non-ignorable. This would be a welcome conclusion for general survey analysis, because it suggests that more complex analysis which allows for non-ignorability is often not essential for adequate modelling of latent attitudes. When this is so, models which do allow for non-ignorable non-response have a primarily diagnostic role, in that they would be fitted to assess whether they need to be included in the final estimation at all. To be able to do this in any given application, we of course still need to be able to estimate models in a framework which also allows for non-ignorability, such as the set of models that is proposed here.

6. Conclusions

When multiple observed items are used as measures of latent variables, it is possible, without very strong additional assumptions, to allow for item non-response which is non-ignorable in that it depends on the latent variables. We have proposed and applied models which implement this idea, building on and extending existing methods of this kind. Our focus has been on the analysis of cross-national surveys and other comparative studies of multiple groups. Investigating item non-response is of particular interest in such studies because the levels and patterns of non-response itself may also vary between countries, and it should be possible to allow for this in the analysis.

The patterns of non-response that are revealed by these models can be of direct interest to survey analysts and designers of surveys. For instance, in our example from the ESS, the probability of answering questions on attitudes towards recipients of welfare benefits was typically higher for men and for middle-aged and more educated individuals. It was also—and perhaps surprisingly—higher for individuals who held more negative levels of the attitude itself, indicating that non-response to these survey questions was often non-ignorable. All these associations were consistent in direction but variable in magnitude and significance across 29 European countries.

The second and often more important purpose of a non-response model is to allow for the non-response in a sufficiently flexible way in the estimation of models for latent attitude variables. Here a non-ignorable non-response model can also be used for sensitivity analysis, to assess whether the results are substantially affected by ignoring the non-response mechanism even when it is really non-ignorable. In our examples, cross-national models for attitudes towards welfare recipients were insensitive in this respect. This would be a reassuring conclusion for general survey analysis. The extent to which it holds in different applications will depend on the levels and patterns of non-response in them.

The model that we have used combines a conventional multigroup latent variable model for the constructs of main interest with a multigroup latent class model for non-response. This also provides an illustration of the flexibility of latent variable modelling more generally, and of how different elements of such models can be of different types and have different roles in the analysis. Comparable models could be formulated also in other contexts where we would want to include models for both substantive constructs and methodological elements of the data in the same analysis.

Acknowledgement

References

Bacci, S. and Bartolucci, F. (2015) A multidimensional finite mixture structural equation model for nonignorable missing responses to test items. *Struct. Equn Modlng*, **22**, 352–365.

Bartholomew, D. J., Knott, M. and Moustaki, I. (2011) *Latent Variable Models and Factor Analysis: a Unified Approach*, 3rd edn. London: Wiley.

Bertoli-Barsotti, L. and Punzo, A. (2013) Rasch analysis for binary data with nonignorable nonresponses. *Psicologica*, **34**, 97–123.

Beunckens, C., Molenberghs, G., Verbeke, G. and Mallincrodt, C. (2008) A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, **64**, 96–105.

Bock, R. D. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29–51.

Boomsma, A. and Hoogland, J. J. (2001) The robustness of LISREL modeling revisited. In *Structural Equation Models: Present and Future—a Festschrift in Honor of Karl Jöreskog* (eds R. Cudeck, S. du Toit and D. Sörbom), pp. 139–168. Lincolnwood: Scientific Software International.

Carpenter, J. R. and Kenward, M. G. (2013) *Multiple Imputation and Its Application*. Chichester: Wiley.

Clogg, C. C. and Goodman, L. A. (1984) Latent structure analysis of a set of multidimensional contingency tables. *J. Am. Statist. Ass.*, **79**, 762–771.

Collins, L. M., Schafer, J. L. and Kam, C.-M. (2001) A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Meth.*, **6**, 330–351.

Cowles, M. C., Carlin, B. P. and Connett, J. E. (1996) Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *J. Am. Statist. Ass.*, **91**, 86–98.

Daniels, M. J. and Hogan, J. W. (2008) *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton: Chapman and Hall.

De Gruttola, V. and Tu, X. M. (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003–1014.

Esping-Andersen, G. (1990) *The Three Worlds of Welfare Capitalism*. Cambridge: Polity.

European Social Survey (2008) *European Social Survey Round 4 Data (2008), Data File Edition 4.2*. Norwegian Social Science Data Services, Bergen.

European Social Survey Data Archive (2009) *Education Upgrade ESS1-ESS4: Documentation Report, Edition 1.2*. (Available from `http://www.europeansocialsurvey.org/docs/methodology/Education_Upgrade_ESS1-4_ED1.2.pdf`.)

Ferragina, E. and Seeleib-Kaiser, M. (2011) Welfare regime debate: past, present, futures? *Poly Polit.*, **39**, 583–611.

Follmann, D. and Wu, M. (1995) An approximate generalized linear model with random effects for informative missing data. *Biometrics*, **51**, 151–168.

Formann, A. K. (2007) Mixture analysis of multivariate categorical data with covariates and missing entries. *Computnl Statist. Data Anal.*, **51**, 5236–5246.

Frangakis, C. E. and Rubin, D. B. (1999) Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, **86**, 365–379.

Glas, C. A. W. and Pimentel, J. L. (2008) Modeling nonignorable missing data in speeded tests. *Educ. Psychol. Measmnt*, **68**, 907–922.

Glas, C. A. W., Pimentel, J. L. and Lamers, S. M. A. (2015) Nonignorable data in IRT models: polytomous responses and response propensity models with covariates. *Psychol. Test Assessmnt Modlng*, **57**, 523–541.

Graham, J. W. (2003) Adding missing-data-relevant variables to FIML-based structural equation models. *Struct. Equn Modlng*, **10**, 80–100.

Hafez, M. S., Moustaki, I. and Kuha, J. (2015) Analysis of multivariate longitudinal data subject to nonrandom dropout. *Struct. Equn Modlng*, **22**, 193–201.

Harel, O. and Schafer, J. L. (2009) Partial and latent ignorability in missing-data problems. *Biometrika*, **96**, 37–50.

Heckman, J. J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econ. Socl Measmnt*, **5**, 475–492.

Hogan, J. W. and Laird, N. M. (1997) Mixture models for the joint distribution of repeated measures and event times. *Statist. Med.*, **16**, 239–257.

Hogan, J. W., Lin, X. and Herman, B. (2004) Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics*, **60**, 854–864.

Holman, R. and Glas, C. A. W. (2005) Modelling non-ignorable missing-data mechanisms with item response theory models. *Br. J. Math. Statist. Psychol.*, **58**, 1–17.

Ibrahim, J. G. and Molenberghs, G. (2009) Missing data methods in longitudinal studies: a review. *Test*, **18**, 1–43.

Jöreskog, K. G. (1971) Simultaneous factor analysis in several populations. *Psychometrika*, **36**, 409–426.

Jung, H., Schafer, J. L. and Seo, B. (2011) A latent class selection model for nonignorably missing data. *Computnl Statist. Data Anal.*, **55**, 802–812.

Kankaraš, M., Moors, G. and Vermunt, J. K. (2011) Testing for measurement invariance with latent class analysis. In *Cross-cultural Analysis: Methods and Applications* (eds E. Davidov, P. Schmidt and J. Billiet), pp. 359–384. New York: Routledge.

Knott, M., Albanese, M. T. and Galbraith, J. (1990) Scoring attitudes to abortion. *Statistician*, **40**, 217–223.

Köhler, C., Pohl, S. and Carstensen, C. H. (2015a) Taking the missing propensity into account when estimating competence scores: evaluation of item response theory models for nonignorable omissions. *Educ. Psychol. Measmnt*, **75**, 850–874.

Köhler, C., Pohl, S. and Carstensen, C. H. (2015b) Investigating mechanisms for missing responses in competence tests. *Psychol. Test Asssessmnt Modlng*, **57**, 499–522.

Korobko, O. B., Glas, C. A. W., Bosker, R. J. and Luyten, J. W. (2008) Comparing the difficulty of examination subjects with item response theory. *J. Educ. Measmnt*, **45**, 137–155.

Lin, H., McCulloch, C. E. and Rosenheck, R. A. (2004) Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics*, **60**, 295–305.

Little, R. J. A. (1995) Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Ass.*, **90**, 1112–1121.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. Hoboken: Wiley.

Lord, F. M. (1974) Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, **39**, 247–264.

Lord, F. M. (1983) Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika*, **48**, 477–482.

Mattei, A. and Ranalli, M. G. (2015) Dealing with non-ignorable nonresponse in survey sampling: a latent modeling approach. *Surv. Methodol.*, **41**, 145–164.

Millsap, R. E. (2011) *Statistical Approaches to Measurement Invariance*. New York: Routledge.

Mislevy, R. J. (2016) Missing responses in item response theory. In *Handbook of Item Response Theory*, vol. 2 (ed. W. J. van der Linden). Boca Raton: Chapman and Hall–CRC.

Mislevy, R. J. and Wu, P. K. (1996) Missing responses and IRT ability estimation: omits, choice, time limits, adaptive testing. *Research Report RR-96-30-ONR*: Educational Testing Service, Princeton.

Molenberghs, G. and Kenward, M. G. (2007) *Missing Data in Clinical Studies*. Hoboken: Wiley.

Moustaki, I. and Knott, M. (2000) Weighting for item non-response in attitude scales by using latent variable models with covariates. *J. R. Statist. Soc.* A, **163**, 445–459.

Moustaki, I. and O'Muircheartaigh, C. (2000) A one dimension latent trait model to infer attitude from non-response for nominal data. *Statistica*, **60**, 259–276.

Muthén, B. O. (2004) *Mplus Statistical Analysis with Latent Variables: Technical Appendices*. Los Angeles: Muthén and Muthén.

Muthén, B., Jo, B. and Brown, H. (2003) Comment on 'Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City'. *J. Am. Statist. Ass.*, **98**, 311–314.

Muthén, B., Kaplan, D. and Hollis, M. (1987) On structural equation modeling with data that are not missing completely at random. *Psychometrika*, **52**, 431–462.

Muthén, L. K. and Muthén, B. O. (2010) *Mplus User's Guide*, 6th edn. Los Angeles: Muthén and Muthén.

Nelson, F. D. (1977) Censored regression models with unobserved, stochastic censoring thresholds. *J. Econmetr.*, **6**, 309–327.

O'Muircheartaigh, C. and Moustaki, I. (1999) Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *J. R. Statist. Soc.*, A, **162**, 177–194.

Pohl, S., Gäfe, L. and Rose, N. (2014) Dealing with omitted and not-reached items in competence tests: evaluating approaches accounting for missing responses in item response theory models. *Educ. Psychol. Measmnt*, **74**, 423–452.

Rose, N. (2013) Item nonresponses in educational and psychological measurement. *PhD Thesis*. Friedrich-Schiller-Universität, Jena.

Rose, N., von Davier, M. and Nagengast, B. (2015) Commonalities and differences in IRT-based methods for nonignorable item nonresponses. *Psychol. Test Assessmnt Modlng*, **57**, 472–498.

Rose, N., von Davier, M. and Xu, X. (2010) Modeling nonignorable missing data with item response theory (IRT). *Research Report RR-10-11*. Educational Testing Service, Princeton.

Roy, J. (2003) Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, **59**, 829–836.

Rubin, D. B. (1976) Inference and missing data (with discussion). *Biometrika*, **63**, 581–592.

Schluchter, M. (1992) Methods for the analysis of informatively censored longitudinal data. *Statist. Med.*, **11**, 1861–1870.

Shao, J. (2007) Handling survey nonresponse in cluster sampling. *Surv. Methodol.*, **33**, 81–85.

Skinner, C. J. and D'Arrigo, J. (2011) Inverse probability weighting for clustered nonresponse. *Biometrika*, **98**, 953–966.

Stoop, I., Billiet, J., Koch, A. and Fitzgerald, R. (2010) *Improving Survey Response: Lessons Learned from the European Social Survey*. Chichester: Wiley.

Svalfors, S., van Oorschot, W., Taylor-Gooby, P., Staerklé, C., Goul Andersen, J., Hills, J. and Mau, S. (2008) Welfare attitudes in a changing Europe. (Available from http://www.europeansocialsurvey.org/docs/round4/questionnaire/ESS4_final_welfare_module_template.pdf.)

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P. and Landis, J. R. (1998) Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, **54**, 367–383.

Wall, M. M., Park, J. Y. and Moustaki, I. (2015) IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Appl. Psychol. Measmnt*, **39**, 583–597.

Wu, M. C. and Bailey, K. R. (1989) Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, **45**, 939–955.

Wu, M. C. and Carroll, R. J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.

Yuan, Y. and Little, R. J. A. (2007) Model-based estimates of the finite population mean for two-stage cluster samples with unit non-response. *Appl. Statist.*, **56**, 79–97.

Yuan, Y. and Little, R. J. A. (2008) Model-based estimates for two-stage cluster samples subject to nonignorable item nonresponse. *J. Off. Statist.*, **23**, 193–211.