# Discrete Choice Nonresponse*

Esmeralda A. Ramalho
CEFAGE-UE
and
Departamento de Economia
Universidade de Évora

Richard J. Smith†
cemmap
and
Faculty of Economics
University of Cambridge

First Draft: January 2002
This Revision: May 2009

**Abstract**

Missing values are endemic in the data sets available to econometricians. This paper suggests a semiparametrically efficient unified likelihood-based approach to deal with general nonignorable missing data problems for discrete choice models. Our concern is when the dependent variable and/or covariates are unobserved for some sampling units. A supplementary random sample of observations on all covariates may be available. A unified treatment of these various sampling structures is presented using a formulation of nonresponse as a modification of choice-based sampling. Extensions appropriate for nonresponse are detailed of Imbens' (1992) efficient generalized method of moments (GMM) estimator for choice-based samples. Simulation evidence reveals very promising results for the various GMM estimators proposed in this paper.

**JEL Classification**: C25, C51.

**Keywords:** Generalized Method of Moments Estimation, Empirical Likelihood, Missing Completely at Random, Nonignorable Nonresponse, Semiparametric Efficiency.

*We are very grateful for the helpful and constructive comments by a Managing Editor and two anonymous referees on an earlier version of this paper. Especial thanks are owed to T. Crossley for discussions concerning sample surveys and S. Hemnavich who kindly advised us concerning the simulation section. Various drafts have been presented at the 2002 Econometric Society European Meeting, Venice, the 2002 European Winter Meetings of the Econometric Society, Budapest, and seminars at U.C. Berkeley, Birkbeck College London, Birmingham, Brown, Cambridge, Carlos III, cemmap, I.F.S. and U.C.L., Erasmus University, Rotterdam, Kent, Leuven, L.S.E., Manchester, Mannheim, M.I.T., Northwestern, Nuffield College, Oxford, University of Pennsylvania, Research Triangle, North Carolina, Southampton, Toulouse, University of Southern California, Warwick and Zurich. We are grateful to participants at those seminars and meetings for their helpful comments. The authors are pleased to acknowledge financial support from respectively Fundação para a Ciência e Tecnologia, program POCTI, partially funded by FEDER, and a 2002 Leverhulme Major Research Fellowship.

†Corresponding Author. Mailing Address: Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, U.K. E-mail Address: rjs27@econ.cam.ac.uk

# 1 Introduction

Survey sampling is principally conducted to gather complete information on all sampling units. Due to a variety of reasons, nonresponse is an unfortunate but endemic feature of sample surveys. Indeed, many surveys now incorporate categories for nonresponse.[1] For a fraction of the subjects either no data at all are available or information on one or more variables is missing. Indeed, some sampling units may simply refuse to participate at all in the study or answer the questionaire incompletely. The interviewer may not be able to contact all the sampling units or fails to ask all questions. Some questionnaires or parts thereof may be destroyed in data processing. Conversely, there are also cases where the presence of missing values is a deliberate part of the sampling process. In variable probability sampling, for example, an observation is randomly drawn from the population and the stratum to which it belongs is identified, the observation being retained in the sample with a probability defined by the agent who collects the sample.[2] Because the latter sampling scheme deliberately generates incomplete data, the mechanism which governs the missingness pattern is known. In the former situation, which is the subject of this paper, in contradistinction, nothing is generally known about the missingness mechanism as data is missing for reasons beyond the control of the researcher.

In econometrics, nonresponse has been addressed in the extensive literature on sample selection pioneered by Heckman (1976) and also in the context of panel data studies, where often some sampling units will drop out after participating in the initial waves of the survey; see, for example, Ridder (1990), Fitzgerald, Gottschalk and Moffitt (1998) and Hirano, Imbens, Ridder and Rubin (2001). In contrast, Horowitz and Manski (1995, 1998, 2001) provide a general discussion of nonparametric identification for regression with missing data on either (both) the variable of interest or (and) the covariates. An enormous statistical literature has also been developed to address the issue of nonresponse; see *inter alia* Little and Rubin (2002) and Schafer (1997). The recent issues, Part 4, 2005, and Part 3, 2006, of *Journal of the Royal Statistical Society* Series A (Statistics in Society) present a number of studies of statistical and econometric interest in which nonresponse features importantly. Texts, see, e.g., Cameron and Trivedi (2005) and Wooldridge (2009), devote sections to consideration of missingness. Two forms of missing data are commonly distinguished: *unit nonresponse* (UNR), where for some sampling units no data at all is available, and *item nonresponse* (INR), where only part of the information is missing. For the former class, much of the literature suggests the use of weighting adjustments, which involve the assignment of weights to respondents to compensate for their systematic differences relative to nonrespondents. For the latter form of nonresponse, many papers propose imputation inference procedures in which the

[1] The Canadian Out-of-Employment Panel Survey allows "Refuse" or "Don't know" as responses in all questions. The British Household Panel Survey records "missing or wild", "inapplicable" or "not answered" for some income related questions.

[2] Moreover, the statistical literature often deals with two-stage sampling designs where in a first stage the main sample is collected and in a second stage further variables, more expensive and/or difficult to collect, are obtained but only for a subset of the survey participants.

missing values are filled in to produce complete data sets.

Many empirical studies, however, do not adopt either of these approaches or that taken in this paper, simply discarding all sampling units with missing values and employing the usual inference procedures associated with random sampling. This practice may seriously bias results when the characteristics of respondents and nonrespondents differ systematically. This nonignorable nature of nonresponse arises because the rate of response may differ across the possible values taken by the dependent variable, i.e., the missing data mechanism is endogenous. Therefore, the observed data provides a distorted picture of the features of the underlying population of interest.

This paper proposes a unified likelihood-based approach for parametric discrete choice models with missing data in a cross-section context. We address a general formulation of missing data which encompasses both INR and UNR, situations where, due to the nature of some of the questions contained in the survey, some sample units either omit the answer to particular questions or refuse to participate in the survey at all. In addition, we allow for the availability of an independent supplementary random sample (SRS) on the covariates. Such information might naturally arise from census data; see, e.g., Cosslett (1981a), Lancaster and Imbens (1996) and Hellerstein and Imbens (1999). Analysis focusses on this general set-up, which may then be specialized for particular missingness schemes including pure INR and UNR and the absence of a SRS. The model appropriate for the available data necessarily becomes an involved function of the underlying structural model for discrete choice and the missing data mechanism. An additional complication typically arises since conditional maximum likelihood estimation is no longer efficient in the presence of nonresponse.

Our approach is semi-parametric. Incomplete data patterns are underpinned by (unknown) missing data mechanisms which are assumed to be completely determined by the discrete outcome variable, i.e., nonresponse is conditionally independent of covariates given the outcome variable. This specification for nonresponse may not be unreasonable in situations where outcomes are associated with social stigma or related to income or wealth and may be of relevance in situations when the outcome variable is latent and only partially observed as illustrated in the simulation section of this paper. Thus, because of the discrete nature of the outcome variable, the probabilities defining nonresponse do not require *a priori* knowledge, but rather are treated as additional parameters to be estimated.[3] To achieve an economy of notation, the main part of the text confines attention to this formulation. Appropriate modifications are described which enable this assumption to be relaxed straightforwardly to permit a degree of discrete dependence on covariates, a formulation which is of especial relevance and interest if some covariates are also discrete. Our approach has the particular advantage that all patterns of missingness may be subsumed in our framework albeit at the expense of a loss of efficiency in circumstances when covariates may only be partially observed. The distribution of the covariates is handled nonparametrically by regarding covariates as discrete, a treatment which parallels that of empirical likelihood and related methods with the covariate discrete probabilities consequently concentrated out on application of maximum likelihood; see *inter*

[3] An alternative approach would specifiy response probabilities parametrically but with the attendant potential for functional form misspecification.

*alia* Owen (2001), Smith (1997, 2001), Newey and Smith (2004) and the references therein. This approach is also adopted in the choice-based (CB) sampling literature, i.e., where covariates are sampled randomly conditional on discrete choice outcomes; see *inter alia* Cosslett (1981a, 1981b, 1993, 1997) and Imbens (1992). Indeed, central to our analysis is a recognition of the similarity between nonresponse and CB samples. Consequently, all of the aforementioned incomplete data patterns may be formulated as modifications of CB sampling. Therefore, application of maximum likelihood (ML) in our context is semi-parametrically efficient and may be regarded as an adaptation and extension of Imbens' (1992) efficient generalized method of moments (GMM) approach for CB sampling; see also Cosslett (1981a, 1981b).[4]

This paper is organized as follows. Section 2 formalizes the model specification for the missing data problems of interest. Section 3 details the observed data likelihoods. GMM estimators are developed and their large sample properties presented in section 4. Specification tests are described in section 5. Extensions to our basic framework are considered in section 6. Section 7 reports some simulation evidence on the performance of some of the proposed estimators based on an application where nonresponse has been considered to be a potentially serious problem. Finally, section 8 concludes. Some technical details are relegated to Appendices A and B. Appendix C details how the missingness mechanism employed in the main body of the paper may be weakened to permit a discrete dependence on covariates.

# 2 Model Specification

## 2.1 Some Notation

Of central concern is the population conditional distribution of the discrete outcome variable $Y$ which takes values in the set $\mathcal{Y} = \{1, ..., C\}$ of $C$ mutually exclusive alternatives given the vector of covariates $X$ with sample space $\mathcal{X}$. The random variable $Y$ and vector $X$ are defined on $\mathcal{Y} \times \mathcal{X}$ with population joint density function

$$f(y, x, \theta) = \mathcal{P}\{y|x, \theta\} f_X(x), \tag{2.1}$$

where the discrete probability $\mathcal{P}\{\cdot|\cdot, \theta\}$ is known up to the $p$-dimensional parameter vector $\theta$. The marginal density function $f_X(\cdot)$ for $X$ is unknown and does not depend on $\theta$. Hence, in the absence of nonresponse, efficient estimation of and inference for $\theta$ would be based on the conditional density $\mathcal{P}\{y|x, \theta\}$. Where there is no loss of clarity, we suppress the dependence on $\theta$ of (2.1) and other joint density functions. The superscript $^0$ is used to denote true values of parameters.

The population marginal probability of observing $Y = y$ is

$$\begin{aligned} Q_y &= \mathcal{P}\{Y = y\} \\ &= \int_{\mathcal{X}} \mathcal{P}\{y|x, \theta\} f_X(x) dx, \end{aligned} \tag{2.2}$$

[4]Subsequent to the preparation of earlier versions and the first submission [Ramalho and Smith (2003)] of this paper, we became aware of Tang *et al.* (2003), which adopts a similar formulation for the missing data mechanism. However, Tang *et al.* (2003) for the discrete choice setting considered here is UNR and is thus a special case of our approach being expressed directly as CB sampling; see, e.g., Cosslett (1981a, b), Imbens (1992) and section 6.1.

where $0 < Q_y < 1$, $y \in \mathcal{Y}$, and $\sum_{v \in \mathcal{Y}} Q_v = 1$. Auxiliary information on the probabilities $Q_y$, $y \in \mathcal{Y}$, may be available or they may in fact be known, e.g., from a large random sample like a census. In the former circumstance, this information is incorporated as in the data combination literature, e.g., Imbens and Lancaster (1994), whereas in the latter case it is treated as exact similarly to the choice-based sampling literature, e.g., Manski and Lerman (1977), Imbens (1992) and Wooldridge (1999, 2001). Cf. section 6.2.

## 2.2 Survey Sampling Structure

The survey objective is to collect a random sample of size $N$ of complete observations on $Y$ and $X$. Suppose, however, that only some sampling units provide all the information requested. These respective samples are designated the *initial* (or *incomplete*) and *complete* samples. Let $n$ denote the number of sampling units which provide information on $Y$.

**Assumption 2.1** *(Initial Sample.) The initial sample is a random sample of size $N$.*

The sample size $N$ is always known for pure INR, since either $Y$ or $X$ are measured for all units. Although the number of unit nonrespondents and, thus, $N$ may not be known to the econometrician, our exposition assumes this knowledge since the analysis is straightforwardly adapted for situations when $N$ is unknown; see section 6.1.[5]

We additionally assume that an independent supplementary random sample (SRS) of observations of size $m$ on $X$ is drawn from the population of interest.

**Assumption 2.2** *(Supplementary Random Sample (SRS).) The SRS of observations of size $m$ on $X$ is independent of the initial sample.*

Let the binary indicator $S$ take value 1 when the sampling unit belongs to the supplementary data set and 0 otherwise. Also define $N_m = N + m$ and $n_m = n + m$. We assume a SRS is always available. If unavailable, the analysis may be adapted by suitably redefining all probabilities given below, suppressing $S = 0$ and replacing $N_m$ by $N$. Alternative $Y = y$ is chosen by $N_y$ individuals, of whom only $n_y$ provide information on $Y$. Hence, $N = \sum_{v \in \mathcal{Y}} N_v$ and $n = \sum_{v \in \mathcal{Y}} n_v$. As all incomplete data problems considered here involve missing data on $Y$, we always observe $n_y$, $n$ and $m$ but never $N_y$, $y \in \mathcal{Y}$.

## 2.3 Missing Data Mechanism

Define the binary indicators

$$I^{\mathcal{Y}} = \begin{cases} 1 & \text{if } Y \text{ is observed} \\ 0 & \text{otherwise} \end{cases} \quad \text{and } I^{\mathcal{X}} = \begin{cases} 1 & \text{if } X \text{ is observed} \\ 0 & \text{otherwise} \end{cases} .$$

[5]Information on $N$ improves inference for the parameters of interest; cf. (3.2) and section 6.1 below. See Li and Qin (1998) for a discussion of several examples of biased data where such information improves semiparametric likelihood-based inference.

Respondent units correspond to the event $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}$ whereas INR and UNR are associated with $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 0\}$ [INR$(y)$], $\{I^{\mathcal{Y}} = 0, I^{\mathcal{X}} = 1\}$ [INR$(x)$] and $\{I^{\mathcal{Y}} = 0, I^{\mathcal{X}} = 0\}$ [UNR] respectively. This descriptive scheme for nonresponse is quite general. Albeit at the expense of a loss of information, sampling units that provide $Y$ and partial information on $X$ may be treated as INR$(y)$ whereas those that provide partial information on $X$ only are treated as UNR.

Conditional independence assumptions are a critical aspect of our analysis and permit a semi-parametric analytical framework to be adopted. Assumptions of this type are quite familiar in econometrics and are a common feature of the treatment effects and nonclassical measurement error literatures; see, e.g., Imbens (2004) and Hu and Schennach (2008). These assumptions, adopted to incorporate general forms of missingness into our analysis, may be relaxed if certain categories of nonresponse are ignored as described below. The nonresponse mechanism is primarily determined by $Y$ which may be empirically reasonable when outcomes are associated with social stigma or related to income or wealth and is of relevance in situations when the outcome variable is latent and only partially observed, see, e.g., section 7.1. Appendix C details how a discrete dependence on covariates may be incorporated. To achieve an economy of notation, however, we confine attention in the main part of the text to missingness mechanisms determined purely in terms of $Y$.

**Assumption 2.3** *(Conditional Probability of Observing $Y$.) Observation of $Y$ is conditionally independent of $X$ given $Y$; i.e.,*

$$\begin{aligned} P_y &= \mathcal{P}\{I^{\mathcal{Y}} = 1 | Y = y, X = x\} \\ &= \mathcal{P}\{I^{\mathcal{Y}} = 1 | Y = y\}, \end{aligned} \tag{2.3}$$

*where $0 < P_y < 1$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$.*

In all cases, we assume that $0 < P_y < 1$. If $P_y = 0$, alternative $Y = y$ would never be observed. If, on the other hand, $P_y = 1$, then there would be no missing values among units with $Y = y$.[6]

**Assumption 2.4** *(Conditional Probability of Observing $X$.) Observation of $X$ is conditionally independent of $X$ given $Y$ and $I^{\mathcal{Y}} = 1$ and of $X$ and $Y$ given $I^{\mathcal{Y}} = 0$; i.e.,*

$$\begin{aligned} G_y &= \mathcal{P}\{I^{\mathcal{X}} = 1 | I^{\mathcal{Y}} = 1, Y = y, X = x\} \\ &= \mathcal{P}\{I^{\mathcal{X}} = 1 | I^{\mathcal{Y}} = 1, Y = y\}, \end{aligned} \tag{2.4}$$

$$\begin{aligned} G^{\mathcal{X}} &= \mathcal{P}\{I^{\mathcal{X}} = 1 | I^{\mathcal{Y}} = 0, Y = y, X = x\} \\ &= \mathcal{P}\{I^{\mathcal{X}} = 1 | I^{\mathcal{Y}} = 0\}, \end{aligned} \tag{2.5}$$

*where $0 < G_y \leq 1$, $0 \leq G^{\mathcal{X}} \leq 1$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$.*

In cases when $G_y = 1$ $X$ is observed for all units that reveal $Y$ and if in addition $G^{\mathcal{X}} = 1$ INR$(x)$ obtains whereas $G_y = 0$ and $G^{\mathcal{X}} = 0$ yields pure UNR.

[6] Identification necessitates the conditional independence of $I^{\mathcal{X}}$ also from $Y$ given $I^{\mathcal{Y}} = 0$ in (2.5).

Figure 1 presents the missingness mechanism structure in the absence of a SRS and summarises the probabilities for the different missingness categories together with the attendant sample sizes defined above together with $n_{yx} = \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{Y}}i_i^{\mathcal{X}}\mathrm{I}(y_i = y)$, $n_x^{nr} = \sum_{i=1}^{N_m}(1-s_i)(1-i_i^{\mathcal{Y}})i_i^{\mathcal{X}}$, and $n_u = \sum_{i=1}^{N_m}(1-s_i)(1-i_i^{\mathcal{Y}})(1-i_i^{\mathcal{X}})$ denoting respectively the numbers of $y$-respondents which provide information on $X$, INR($x$) units, and UNR units. The event hierarchy adopted in Assumptions 2.3 and 2.4 and illustrated in Figure 1 is general, other possible structures being observationally equivalent.

**Figure 1 about here**

By Assumptions 2.2 and 2.3, $\mathcal{P}\{I^{\mathcal{Y}} = 1|Y = y, X = x, S = 0\} = \mathcal{P}\{I^{\mathcal{Y}} = 1|Y = y\} = P_y$. Furthermore, from Assumptions 2.2, 2.3, and 2.4, $\mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1|Y = y, X = x, S = 0\} = P_yG_y$ etc., i.e., conditional probabilities of observing respondent, INR($y$), INR($x$) or UNR units given $(Y, X)$ although dependent on $Y$ are independent of $X$.

Aspects of Assumptions 2.3 and 2.4 may be dispensed with if observations on INR($y$), INR($x$) or UNR units are ignored or are unavailable. E.g., if consideration is confined to respondent and UNR units only, Assumption 2.4 may be dropped by redefining probabilities conditional on $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}$ and $\{I^{\mathcal{Y}} = 0, I^{\mathcal{X}} = 0\}$, i.e., $P_y = \mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1|Y = y, X = x, \{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}, \{I^{\mathcal{Y}} = 0, I^{\mathcal{X}} = 0\}\}$, $y \in \mathcal{Y}$, and, in particular, $G_y = 1$, $y \in \mathcal{Y}$, and $G^{\mathcal{X}} = 0$, cf. (2.3), (2.4) and (2.5). See section 6.1 for related discussion.

Combining (2.2), (2.3) and (2.4), the probability of observing a respondent unit is

$$\mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\} = \sum_{v\in\mathcal{Y}} P_vG_vQ_v, \tag{2.6}$$

which, because, in general, $P_y$ and $G_y$ are unknown, will also be unknown even if $Q_y$, $y \in \mathcal{Y}$, are known.

If the rate of response is the same for all alternatives, i.e., $P_y = P$ and $G_y = G$, $y \in \mathcal{Y}$, data are *missing completely at random* (MCAR) since $(I^{\mathcal{Y}}, I^{\mathcal{X}})$ is independent of $(Y, X)$; see Little and Rubin (2002, p.12). Note that, although the missingness mechanism is ignorable, INR($y$) units are not; see section 5.1. Ignorable nonresponse requires the additional condition $G = 1$, in which case the complete sample is also random. If, however, INR($y$) units are discarded and probabilities (2.3), (2.4) and (2.5) consequently redefined conditional on the complement of $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 0\}$, then $G_y = 1$, $y \in \mathcal{Y}$, and nonresponse *is* ignorable if data MCAR. Random sample ML applied to the complete sample $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}$ units may then be used.[7]

[7]If only information on $X$ was missing, $I^{\mathcal{Y}} = 1$ and, thus, $P_y = 1$, $y \in \mathcal{Y}$, i.e., pure INR($y$). Hence, according to the mechanism (2.4), data would be *missing at random* (MAR), since the probability of recording $X$ is independent of $X$ after controlling for $Y$; see Rubin (1976) and Little and Rubin (2002). The missingness mechanism is thus ignorable for likelihood-based inference; cf. (3.3). Most of the statistical literature on nonresponse focusses on data MAR, dealing mainly with procedures for imputing missing values; see, e.g., Little and Rubin (2002) and Schafer (1997).

### 2.4 Missing Data Formulation by Stratification

An important point of departure for this paper is the adaptation of the approach taken in the CB sampling literature to the missing data problems considered here. In order to do so, we reinterpret the different forms of response and nonresponse as strata for each discrete value of $Y$. A further stratum including SRS units is added. The proportions of each of the $C$ $y$-respondent strata, i.e., $I^{\mathcal{Y}} = 1$, in the sample and in the population are denoted by $H_y$ and $Q_y$ respectively; see (2.2). Each of the $C$ $y$-nonrespondent strata, i.e., $I^{\mathcal{Y}} = 0$, has sampling proportion $H_y^{nr}$ but the same population proportion $Q_y$. Therefore, the initial random sample is interpreted as a combination of two CB samples consisting of respondent and nonrespondent sampling units. Finally, the SRS stratum has a proportion of $\mathcal{P}\{S = 1\} = H_s$ in the sample, while in the population, as the supplementary sample is random, we observe units from this stratum with probability 1.

The probability of observing a $y$-respondent unit and $Y = y$ is

$$H_y = \mathcal{P}\{I^{\mathcal{Y}} = 1, Y = y, S = 0\}, \tag{2.7}$$

while the corresponding probability for $y$-nonrespondent units is

$$H_y^{nr} = \mathcal{P}\{I^{\mathcal{Y}} = 0, Y = y, S = 0\}. \tag{2.8}$$

Aggregating over $\mathcal{Y}$ yields the probability of observing, respectively, $y$-respondent, $\mathcal{P}\{I^{\mathcal{Y}} = 1, S = 0\} = \sum_{v\in\mathcal{Y}} H_v$, and $y$-nonrespondent units, $\mathcal{P}\{I^{\mathcal{Y}} = 0, S = 0\} = \sum_{v\in\mathcal{Y}} H_v^{nr}$. Hence,

$$\begin{aligned} \mathcal{P}\{S &= 0\} = 1 - H_s \\ &= \sum_{v\in\mathcal{Y}} H_v + \sum_{v\in\mathcal{Y}} H_v^{nr}. \end{aligned}$$

From Assumption 2.2, by independence, the marginal population probability $Q_y = \mathcal{P}\{Y = y|S = 0\}$. Thus, as $\mathcal{P}\{Y = y, S = 0\} = \sum_{i=0}^{1} \mathcal{P}\{I^{\mathcal{Y}} = i, Y = y, S = 0\}$,

$$Q_y(1 - H_s) = H_y + H_y^{nr}, \tag{2.9}$$

which permits the elimination of the unknown sample probabilities $H_y^{nr}$, $y \in \mathcal{Y}$. Also, by Assumptions 2.2 and 2.3, cf. (2.9), since $P_y = \mathcal{P}\{I^{\mathcal{Y}} = 1|Y = y, S = 0\}$,

$$P_y = \frac{H_y}{Q_y\,(1 - H_s)}. \tag{2.10}$$

From (2.10), as $0 < P_y < 1$ by Assumption 2.3, $0 < H_y < Q_y\,(1 - H_s)$. In all cases $H_y$ may be estimated from the incomplete sample as $n_y/N_m$. Hence, (2.10) may be used to estimate $P_y$ when $Q_y$ is either known or estimated by the methods set out in section 4.

## 3 Observed Data Likelihoods

This section considers the likelihood function for the observed data, as well as other sampling densities of interest. We use the generic notation $h(\cdot)$ for sample density functions.

The joint sample density function for $Y$, $X$, $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$ is

$$\begin{aligned}
h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \Bigg\{\Big[h(y,x,i^{\mathcal{Y}}=1,i^{\mathcal{X}}=1,s=0)^{i^{\mathcal{X}}}h(y,i^{\mathcal{Y}}=1,i^{\mathcal{X}}=0,s=0)^{(1-i^{\mathcal{X}})}\Big]^{i^{\mathcal{Y}}} \\
&\times \Big[h(x,i^{\mathcal{Y}}=0,i^{\mathcal{X}}=1,s=0)^{i^{\mathcal{X}}}h(i^{\mathcal{Y}}=0,i^{\mathcal{X}}=0,s=0)^{(1-i^{\mathcal{X}})}\Big]^{(1-i^{\mathcal{Y}})}\Bigg\}^{1-s} \\
&\times h\left(x,s=1\right)^{s} \\
&= \Bigg\{\Big[(\mathcal{P}\{y,I^{\mathcal{Y}}=1,I^{\mathcal{X}}=1,S=0\}h(x|y))^{i^{\mathcal{X}}}(\mathcal{P}\{y,I^{\mathcal{Y}}=1,I^{\mathcal{X}}=0,S=0\})^{(1-i^{\mathcal{X}})}\Big]^{i^{\mathcal{Y}}} \\
&\times \Bigg[(\sum_{v\in\mathcal{Y}}\mathcal{P}\{v,I^{\mathcal{Y}}=0,I^{\mathcal{X}}=1,S=0,\}h(x|v))^{i^{\mathcal{X}}} \\
&\times (\sum_{v\in\mathcal{Y}}\mathcal{P}\{v,I^{\mathcal{Y}}=0,I^{\mathcal{X}}=1,S=0,\})^{(1-i^{\mathcal{X}})}\Bigg]^{(1-i^{\mathcal{Y}})}\Bigg\}^{1-s}[H_s f_X(x)]^s. \qquad (3.1)
\end{aligned}$$

The second equality in (3.1) arises since $h(x|y,i^{\mathcal{Y}},i^{\mathcal{X}},S=0)=h(x|y)$ because, from Assumption 2.2, $h\left(x|y,i^{\mathcal{Y}},i^{\mathcal{X}},S=0\right)=h\left(x|y,i^{\mathcal{Y}},i^{\mathcal{X}}\right)$ and $h\left(x|y,i^{\mathcal{Y}},i^{\mathcal{X}}\right)=h\left(x|y\right)$ by Assumptions 2.3 and 2.4, the latter equality paralleling CB sampling. Therefore, eliminating the dependence on the unknown probabilities $H_y^{nr}$ using (2.9),

$$\begin{aligned}
h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \Bigg\{\Bigg[\left(\frac{H_y}{Q_y}G_y\mathcal{P}\{y|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}} \\
&\times \left(\frac{H_y}{Q_y}(1-G_y)\int_{\mathcal{X}}\mathcal{P}\{y|x,\theta\}f_X(x)dx\right)^{(1-i^{\mathcal{X}})}\Bigg]^{i^{\mathcal{Y}}} \\
&\times \Bigg[\left(\sum_{v\in\mathcal{Y}}(1-H_s-\frac{H_v}{Q_v})G^{\mathcal{X}}\mathcal{P}\{v|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}} \\
&\times \left(\sum_{v\in\mathcal{Y}}(1-H_s-\frac{H_v}{Q_v})(1-G^{\mathcal{X}})\int_{\mathcal{X}}\mathcal{P}\{y|x,\theta\}f_X(x)dx\right)^{(1-i^{\mathcal{X}})}\Bigg]^{(1-i^{\mathcal{Y}})}\Bigg\}^{1-s} \\
&\times [H_s f_X(x)]^s \\
&= \Bigg\{\Bigg[\left(\frac{H_y}{Q_y}G_y\mathcal{P}\{y|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}}(H_y(1-G_y))^{(1-i^{\mathcal{X}})}\Bigg]^{i^{\mathcal{Y}}} \qquad (3.2) \\
&\times \Bigg[\left((1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\})G^{\mathcal{X}}f_X(x)\right)^{i^{\mathcal{X}}} \\
&\times \left((1-H_s-\sum_{v\in\mathcal{Y}}H_v)(1-G^{\mathcal{X}})\right)^{(1-i^{\mathcal{X}})}\Bigg]^{(1-i^{\mathcal{Y}})}\Bigg\}^{1-s}[H_s f_X(x)]^s.
\end{aligned}$$

The constituent of (3.2) associated with the joint indicator $(1-S)I^{\mathcal{Y}}I^{\mathcal{X}}$ contains information provided by respondent units and corresponds to the complete data density. Crucially, this component fundamentally differs from the population joint density function (2.1) of $Y$ and $X$ which would be appropriate under random sampling. Hence, unless the data are MCAR, in which case $H_y/Q_y$ and $G_y$ are invariant and thus irrelevant for likelihood-based inference, and $G_y = 1$, $y \in \mathcal{Y}$, random sample procedures should not be used with the complete sample; see section 5.1. The second, third and fourth terms in (3.2) detail information provided by INR($y$), INR($x$) and UNR nonrespondent units, the second and third terms containing additional information on the discrete outcome variable and covariates respectively with the fourth term merely incorporating information on the total sample size $N$ which is employed in the estimation of $H_y$, $H_s$ and $G^{\mathcal{X}}$. The final component of (3.2) is information on $X$ provided by individuals in the SRS. Note that the same data on respondent and SRS units is observed for all missingness patterns.

The joint sample likelihood (3.2) also allows the incorporation of other nonrespondent data structures. If $X$ is only partially observed, then such sample units may be treated as either INR($y$) or UNR at the expense of an attendant loss of information and consequent estimator inefficiency. Note that the data on $X$ reported by INR($x$) and SRS units enter (3.2) in quite different ways.

### 3.1 Unit and Item Nonresponse

The sample density function (3.2) is straightforwardly specialised to pure INR by setting $G^{\mathcal{X}} = 1$ yielding the joint sample density function for $Y$, $X$, $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$ as

$$\begin{aligned} h_{INR}(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \left\{\left[\left(\frac{H_y}{Q_y}G_y\mathcal{P}\{y|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}}(H_y(1-G_y))^{(1-i^{\mathcal{X}})}\right]^{i^{\mathcal{Y}}}\right. \\ &\quad \left.\times\left((1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\})f_X(x)\right)^{i^{\mathcal{X}}(1-i^{\mathcal{Y}})}\right\}^{1-s}[H_sf_X(x)]^s. \end{aligned} \tag{3.3}$$

Pure INR($x$) (INR($y$)) nonresponse is obtained if $G_y = 1$ ($P_y = 1$), $y \in \mathcal{Y}$, and, since $I^{\mathcal{X}} = 1$ ($I^{\mathcal{Y}} = 1$), terms indexed by $1 - I^{\mathcal{X}}$ ($1 - I^{\mathcal{Y}}$) are suppressed.

Correspondingly, for pure UNR, setting $G_y = 1$ and $G^{\mathcal{X}} = 0$ in (3.2), thus suppressing terms indexed by $I^{\mathcal{X}}\left(1-I^{\mathcal{Y}}\right)$ and $I^{\mathcal{Y}}\left(1-I^{\mathcal{X}}\right)$,

$$\begin{aligned} h_{UNR}(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \left\{\left(\frac{H_y}{Q_y}\mathcal{P}\{y|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}i^{\mathcal{Y}}}\left(1-H_s-\sum_{v\in\mathcal{Y}}H_v\right)^{(1-i^{\mathcal{X}})(1-i^{\mathcal{Y}})}\right\}^{1-s} \\ &\quad \times[H_sf_X(x)]^s. \end{aligned} \tag{3.4}$$

### 3.2 Ancillarity

Efficient likelihood-based inference may be conducted by conditioning on weakly exogenous or, equivalently, (partially) ancillary statistics, see Engle, Hendry and Richard (1983) and Basu (1977).

### 3.2.1 Density function of $X$

In contrast to the population density function (2.1), the covariates $X$ are in general not ancillary for $\theta$, i.e., the sampling density $h_X(x)$ of $X$ obtained from (3.2) is functionally dependent on $\theta$; *viz.*

$$\begin{aligned} h_X(x) &= \sum_{s=0}^{1}\sum_{i^{\mathcal{Y}}=0}^{1}\sum_{i^{\mathcal{X}}=0}^{1}\sum_{v\in\mathcal{Y}} h(v,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) \\ &= f_X(x)\left((1-H_s)G^{\mathcal{X}} - \sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}(G^{\mathcal{X}}-G_v)\mathcal{P}\{v|x,\theta\} + H_s\right) \\ &\quad +(1-H_s)(1-G^{\mathcal{X}}) + \sum_{v\in\mathcal{Y}} H_v(G^{\mathcal{X}}-G_v). \end{aligned} \tag{3.5}$$

Conditional ML given $X$ is therefore inefficient and estimation should be based on (3.2).[8] For pure INR($x$), however, $h_X(\cdot)$ reduces to the population density function $f_X(\cdot)$. In this case, as the density $f_X(\cdot)$ factors out, efficient inference for $\theta$ can be conducted based on the conditional density given $X$, suppressing the index $I^{\mathcal{X}}$,

$$h_{INR(x)}(y,i^{\mathcal{Y}},s|x) = \left\{\left(\frac{H_y}{Q_y}\mathcal{P}\{y|x,\theta\}\right)^{i^{\mathcal{Y}}}\left((1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\})\right)^{(1-i^{\mathcal{Y}})}\right\}^{1-s}[H_s]^s. \tag{3.6}$$

### 3.2.2 Joint Density of $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$

In contradistinction to that for $X$, the joint density of the indicators $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$ does not depend on $\theta$, *viz.*,

$$\begin{aligned} h(i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \sum_{v\in\mathcal{Y}}\int_{\mathcal{X}} h(v,x,i^{\mathcal{Y}},i^{\mathcal{X}},s)dx \\ &= \left\{\left[\left(\sum_{v\in\mathcal{Y}}H_vG_v\right)^{i^{\mathcal{X}}}\left(\sum_{v\in\mathcal{Y}}H_v(1-G_v)\right)^{(1-i^{\mathcal{X}})}\right]^{i^{\mathcal{Y}}}\right. \\ &\quad \times\left[\left((1-H_s-\sum_{v\in\mathcal{Y}}H_v)G^{\mathcal{X}}\right)^{i^{\mathcal{X}}}\right. \\ &\quad \left.\left.\times\left((1-H_s-\sum_{v\in\mathcal{Y}}H_v)(1-G^{\mathcal{X}})\right)^{(1-i^{\mathcal{X}})}\right]^{(1-i^{\mathcal{Y}})}\right\}^{1-s}[H_s]^s. \end{aligned} \tag{3.7}$$

However, as the conditional density for $Y$ and $X$ given $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$ obtained from (3.2) and (3.7) also involves the parameters $H_y$, $G_y$, $G^{\mathcal{X}}$ and $H_s$, $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$ and $S$ are not ancillary for $\theta$. Therefore, similarly to Imbens (1992) and Imbens and Lancaster (1996) for endogenous stratified sampling, estimation is based on the unconditional likelihood function (3.2).

[8] For a discussion on the issue of covariate ancillarity for problems when data MAR, see Lawless, Kalbfleisch and Wild (1999).

# 4 Efficient Generalized Method of Moments

This section considers ML estimation applied to the (unconditional) log-likelihood function based on (3.2). The marginal distribution $f_X(\cdot)$ of $X$ is treated semiparametrically by defining the covariate sample space $\mathcal{X}$ as if it is discrete with associated probability masses defined by $\mathcal{P}\{X = x\} = \pi_x$, $0 < \pi_x < 1$, $x \in \mathcal{X}$.[9] From an efficiency standpoint, this analytical device is innocuous; see Theorem 4.2 and Appendix B. The nuisance parameters $\pi_x$, $x \in \mathcal{X}$, may be concentrated out as demonstrated in Appendix A resulting in a set of moment indicators which represent an adaptation and extension to the missing data context of the basis of the efficient GMM estimation method developed by Imbens (1992) for CB sampling.[10] The marginal population stratum probabilities $Q_y$, $y \in \mathcal{Y}$, may be either unknown or auxiliary information may be available as in section 6.2. Our reinterpretation of incomplete data problems for discrete choice models using a CB sampling setting suggests that some of the estimators originally proposed for that set-up may be relevant here also. In particular, as noted in section 6.1, all CB sampling estimators may be used to deal with UNR when the initial sample size $N$ is unknown.

The remainder of this section is organized as follows. Section 4.1 derives the moment indicators appropriate for handling all the missing data patterns considered in this paper. Section 4.2 details the large sample properties of the resultant GMM estimator.

## 4.1 Moment Indicators

The (unconditional) log-likelihood function based on (3.2) is

$$\begin{aligned}\log L \;=\; \sum_{i=1}^{N_m} \Bigg\{ & (1-s_i)\, i_i^{\mathcal{Y}} i_i^{\mathcal{X}} \log\left(\frac{H_{y_i}}{Q_{y_i}} G_{y_i} \mathcal{P}\{y_i|x_i,\theta\}\pi_{x_i}\right) \\ & + (1-s_i)\, i_i^{\mathcal{Y}}(1-i_i^{\mathcal{X}}) \log\left(H_{y_i}(1-G_{y_i})\right) \\ & + (1-s_i)\,(1-i_i^{\mathcal{Y}}) i_i^{\mathcal{X}} \log\left((1-H_s-\sum_{v\in\mathcal{Y}} \frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}) G^{\mathcal{X}} \pi_{x_i}\right) \\ & + (1-s_i)\,(1-i_i^{\mathcal{Y}})(1-i_i^{\mathcal{X}}) \log\left((1-H_s-\sum_{v\in\mathcal{Y}} H_v)(1-G_i^{\mathcal{X}})\right) + s_i\left(\log H_s + \log \pi_{x_i}\right) \Bigg\},\end{aligned} \tag{4.1}$$

where $Q_y = \sum_{x\in\mathcal{X}} \pi_x \mathcal{P}\{y|x,\theta\}$, $y \in \mathcal{Y}$. Maximization of (4.1) is undertaken subject to the restriction $\sum_{x\in\mathcal{X}} \pi_x = 1$.

[9] If $\mathcal{X}$ is assumed to consist of each observation on $X$ then this approach directly parallels that of Cosslett (1981a, 1981b, 1993, 1997) and is equivalent to one based on empirical likelihood [Owen (2001)]. Chamberlain (1987) uses this method to deduce the semi-parametric efficiency lower bounds for both unconditional and conditional moment restriction models in a random sampling setting.

[10] Although our discussion emphasises GMM, other asymptotically equivalent methods such as general empirical likelihood (GEL) [Smith (1997, 2001) and Newey and Smith (2004)] are applicable. GEL includes empirical likelihood [Qin and Lawless (1994), Imbens (1997) and Owen (2001)], exponential tilting [Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)] and the continuous updating estimator [Hansen, Heaton, and Yaron (1996)] as special cases. GMM and GEL estimators are identical for the just-identified context of this section; cf. section 6.2.

From Appendix A, (A.8), (A.9) and (A.10), the system of GMM moment indicators is

$$
\begin{aligned}
H_t \;:\; & (1-s)i^{\mathcal{Y}}\frac{1}{H_t}\mathrm{I}(y=t)-(1-s)(1-i^{\mathcal{Y}})(1-i^{\mathcal{X}})\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v} \\
& -(1-s)(1-i^{\mathcal{Y}})i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}\frac{\mathcal{P}\{t|x,\theta\}}{Q_t},\ t\in\mathcal{Y}, \qquad (4.2)\\
G_t \;:\; & (1-s)i^{\mathcal{Y}}(i^{\mathcal{X}}-G_t)\mathrm{I}(y=t),\ t\in\mathcal{Y}, \qquad (4.3)\\
G^{\mathcal{X}} \;:\; & (1-s)(1-i^{\mathcal{Y}})(i^{\mathcal{X}}-G^{\mathcal{X}}), \qquad (4.4)\\
H_s \;:\; & s-H_s, \qquad (4.5)\\
\theta \;:\; & (1-s)i^{\mathcal{Y}}i^{\mathcal{X}}\frac{\partial\log\mathcal{P}\{y|x,\theta\}}{\partial\theta} \qquad (4.6)\\
& +((1-s)i^{\mathcal{X}}+s) \\
& \times\left(1-(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_v})(1-G^{\mathcal{X}})\frac{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right. \\
& \left.-\sum_{v\in\mathcal{Y}}O_v\frac{\mathcal{P}\{v|x_i,\theta\}}{Q_v}\right)^{-1}\sum_{v\in\mathcal{Y}}\left(O_v-\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right. \\
& \left.\times(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_v})(1-G^{\mathcal{X}})H_v\right)\frac{1}{Q_v}\frac{\partial\mathcal{P}\{v|x,\theta\}}{\partial\theta} \\
& -(1-s)(1-i^{\mathcal{Y}})i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\frac{\partial\mathcal{P}\{v|x,\theta\}}{\partial\theta}\Bigg], \\
Q_y \;:\; & Q_y-((1-s)i^{\mathcal{X}}+s) \qquad (4.7)\\
& \times\left(1-(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_v})(1-G^{\mathcal{X}})\frac{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right. \\
& \left.-\sum_{v\in\mathcal{Y}}O_v\frac{\mathcal{P}\{v|x_i,\theta\}}{Q_v}\right)^{-1}\mathcal{P}\{y|x,\theta\}, \\
O_t \;:\; & (1-s)i^{\mathcal{Y}}(1-i^{\mathcal{X}})\mathrm{I}(y=t)-O_t,\ t\in\mathcal{Y}, \qquad (4.8)
\end{aligned}
$$

where $\mathrm{I}(\cdot)$ denotes an indicator function. The system of moment indicators (4.2)-(4.8) incorporates the additional parameters $O_y$, $y\in\mathcal{Y}$, alongside $Q_y$, $y\in\mathcal{Y}$, with associated moment indicators (4.8) and (4.7). The presence of the additional terms in (4.2) reflects the additional information conveyed by the covariate information from nonrespondents for the stratum probabilities, $H_y$, over and above that of the sample proportions, $n_y/N_m$, $y\in\mathcal{Y}$. The first component in (4.6) is the score vector associated with random sample ML based on the complete sample. The other terms effectively mean-adjust this moment indicator vector to achieve consistent parameter estimation.[11]

[11]The ML estimators $\hat{G}_y=n_{yx}/n_y$, $\hat{G}^{\mathcal{X}}=n_x^{nr}/(N-n)$ and $H_s=m/N_m$ for $G_y$, $G^{\mathcal{X}}$ and $H_s$ from (4.1) are also the marginal ML estimators obtained from (3.7). The derivation of the moment indicators (4.2)-(4.8) from (A.8), (A.9) and (A.10) makes use of the identities $n_u=(N-n)(1-\hat{G}^{\mathcal{X}})$, $N-n=N_m(1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{O}_v/(1-\hat{G}_v))$, $n_y-n_{yx}=n_y(1-\hat{G}_y)$ and $N_m\hat{O}_y=n_y-n_{yx}$.

For some missingness patterns, the moment indicator system (4.2)-(4.8) substantially simplifies with some terms and indexes being eliminated; cf. section 3.1. Additionally, some parameter moment indicators are also completely suppressed, e.g., for pure INR($x$), $G_y$, $y \in \mathcal{Y}$, and $G^{\mathcal{X}}$ and for UNR, $G_y$, $y \in \mathcal{Y}$, $G^{\mathcal{X}}$ and $O_y$, $y \in \mathcal{Y}$. In the latter case, $\hat{H}_y = \hat{O}_y = n_y/N_m$, $y \in \mathcal{Y}$, and the moment indicator (4.2) for $H_t$ simplifies to $(1-s)i^{\mathcal{Y}}\mathrm{I}(y=t) - H_t,\ t \in \mathcal{Y}$.

In general, unless data are MCAR and $G_y = 1$, $y \in \mathcal{Y}$, and, thus, the second and third terms in (4.6) vanish, conventional random sample estimators applied to the complete data set are inconsistent. However, Carroll, Ruppert and Stefanski (1995, p.184) and Allison (2001, p.7), aver that as long as $\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}$ is conditionally independent of $X$ given $Y$, i.e., the missingness mechanism Assumptions 2.3 and 2.4, then estimators for the slope parameters of logit models remain consistent in apparent contradiction to the results presented here. As shown in Ramalho and Smith (2003, section 4.4), their conjecture results from the particular properties of the multiplicative intercept model (MIM) class, which includes the logit model as a particular case and is also widely discussed in the CB sampling area. The CB sampling literature demonstrates that both intercept terms and marginal choice probabilities $Q_y$ are not separately identified in MIM when these probabilities are unknown. Moreover, except for the shift in intercept terms, all parameters in MIM are consistently estimated by random sample ML; see, e.g., Hsieh, Manski and McFadden (1985) and Weinberg and Wacholder (1993). UNR preserves these two characteristics. However, neither of these properties can be extended to the general case considered here unless incomplete units are discarded which again reduces to UNR. See Ramalho and Smith (2003, section 4.4) for a detailed discussion of these points. If one wishes to include the additional information from nonrespondents and/or the SRS in the estimation procedure, then the GMM estimator proposed here is appropriate.

### 4.2 GMM Estimation

Let $\varphi$ denote $H_y$, $G_y$, $y \in \mathcal{Y}$, $G^{\mathcal{X}}$, $H_s$ and $\theta$ together with $Q_y$ and $O_y$, $y \in \mathcal{Y}$, and $\varphi^0$ the true value of $\varphi$. Define $g(\varphi)$ as the vector of moment indicators obtained after stacking (4.2)-(4.8). A subscript $i$ denotes evaluation at observation $(y_i, x_i, i_i^{\mathcal{Y}}, i_i^{\mathcal{X}}, s_i)$, $(i = 1, ..., N_m)$.

The GMM objective function is defined by

$$\hat{J}(\varphi) = \hat{g}(\varphi)'\hat{W}\hat{g}(\varphi), \tag{4.9}$$

where $\hat{W}$ is a positive semi-definite weighting matrix. The vector $\hat{g}(\varphi) = \sum_{i=1}^{N_m} g_i(\varphi)/N_m$ is the sample counterpart of the moment conditions $E\left[g(\varphi^0)\right] = 0$, where $E[\cdot]$ denotes expectation taken over $h(y, x, i^{\mathcal{Y}}, i^{\mathcal{X}}, s)$ of (3.2). Let $\hat{\varphi}$ denote the minimiser of (4.9).

We invoke the following standard regularity conditions which are sufficient for the consistency and asymptotic normality of $\hat{\varphi}$. See Imbens (1992) and Newey and McFadden (1994, Theorems 2.6 and 3.4).

**Assumption 4.1** *(a)* $\theta^0 \in int(\Theta)$, $\Theta$ *a compact subset of* $\mathcal{R}^p$*; (b)* $H_y^0 > 0$, $G_y^0 > 0$, $y \in \mathcal{Y}$, $(G^{\mathcal{X}})^0 > 0$ *and* $H_s^0 > 0$.

**Assumption 4.2** *(a) $\mathcal{P}\{y|x,\theta\}$ is twice continuously differentiable in $\theta \in \Theta$; (b) $\mathcal{P}\{y|x,\theta\}$ and $\partial\mathcal{P}\{y|x,\theta\}/\partial\theta$ are continuous at each $\theta \in \Theta$; (c) $\mathcal{P}\{y|x,\theta\} > 0$, $y \in \mathcal{Y}$, for all $x \in \mathcal{X}$ and $\theta$ in an open neighbourhood of $\theta^0$; (d) $f_X(x) > 0$ for all $x \in \mathcal{X}$; (e) $1 - H_s > \sum_{v\in\mathcal{Y}}(H_v/Q_v)\mathcal{P}\{v|x,\theta\}$ for all $x \in \mathcal{X}$ and $\varphi$ in an open neighbourhood of $\varphi^0$.*

Assumptions 4.2 (c) and (d) ensure that $Q_y^0 > 0$, $y \in \mathcal{Y}$. Assumption 4.2 (e) requires a positive sample (and population) probability of observing $\{I^{\mathcal{Y}} = 0, S = 0\}$, an assumption which is not required for UNR when $N$ is unknown.

Let $G^0 = E\left[\partial g(\varphi^0)/\partial\varphi'\right]$ and $\Omega^0 = E\left[g(\varphi^0)g(\varphi^0)'\right]$. Note that the parameter vector $\varphi^0$ is just-identified.

**Assumption 4.3** *(a) $\hat{W} \xrightarrow{p} W$, $W$ positive definite; (b) $\varphi^0$ is the unique solution to $E[g(\varphi^0)] = 0$; (c) $E[\sup_\varphi \|g(\varphi)\|^2] < \infty$ and $E[\sup_{\varphi\in\mathcal{N}} \|\partial g(\varphi)/\partial\varphi'\|] < \infty$ where $\mathcal{N}$ is a neighbourhood of $\varphi^0$; (d) $\Omega^0$ is nonsingular; (e) $G^0$ is full column rank.*

These conditions lead to the following result.

**Theorem 4.1** *(Consistency and Asymptotic Normality of $\hat{\varphi}$.) If Assumptions 2.1-2.4 and 4.1-4.3 are satisfied then*

$$\hat{\varphi} \xrightarrow{p} \varphi^0, N_m^{1/2}\left(\hat{\varphi} - \varphi^0\right) \xrightarrow{d} N(0, (G^0)^{-1}\Omega^0(G^0)'^{-1}), \tag{4.10}$$

*where $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution respectively.*

When $X$ is discrete, $\hat{\varphi}$ is the ML estimator for $\varphi^0$ and is, thus, asymptotically first order efficient. Asymptotic efficiency, in the semiparametric sense, is proved for the general case analogously to Theorem 3.3 in Imbens (1992); see Appendix B.

**Theorem 4.2** *(Semiparametric Efficiency of $\hat{\varphi}$.) If Assumptions 2.1-2.4 and 4.1-4.3 are satisfied then $\hat{\varphi}$ achieves the semiparametric efficiency bound.*

# 5 Specification Tests

## 5.1 Missing Completely At Random

In practice, whether the missingness mechanism is ignorable would be unknown. If information on the population probabilities $Q_y$ were available, a comparison of $Q_y$ with the sampling proportion $H_y$ might be used to draw rough conclusions about the nature of the missing data. More formally, specification tests for the null hypothesis of data MCAR may be constructed as described below.

If the data are MCAR, $P_y$ and $G_y$ are constant for all $y \in \mathcal{Y}$, i.e., $P_y = P$ and $G_y = G$, $y \in \mathcal{Y}$; see below (2.6). From (2.10), the MCAR null hypothesis is

$$\frac{H_y}{Q_y} = P(1 - H_s),\ G_y = G, y \in \mathcal{Y}. \tag{5.1}$$

GMM estimation under (5.1) using the moment indicator systems (4.2)-(4.8) is straightforward. From (3.2), the joint sample density becomes

$$\begin{aligned} h^{MCAR}(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) &= \left[\mathcal{P}\{y|x,\theta\}^{i^{\mathcal{X}}}(Q_y)^{1-i^{\mathcal{X}}}\right]^{(1-s)i^{\mathcal{Y}}}\left[P^{i^{\mathcal{Y}}}(1-P)^{1-i^{\mathcal{Y}}}\right]^{1-s} \\ &\quad\times\left[G^{i^{\mathcal{X}}}(1-G)^{1-i^{\mathcal{X}}}\right]^{(1-s)i^{\mathcal{Y}}}\left[(G^{\mathcal{X}})^{i^{\mathcal{X}}}(1-G^{\mathcal{X}})^{1-i^{\mathcal{X}}}\right]^{(1-s)(1-i^{\mathcal{Y}})} \\ &\quad\times\left[H_s^s(1-H_s)^{1-s}\right]f_X(x)^{(1-s)i^{\mathcal{X}}+s}. \end{aligned}$$

Therefore, the MCAR estimators are $\tilde{P}=n/N$, $\tilde{G}=\sum_{y\in\mathcal{Y}}n_{yx}/n$, $\tilde{G}^{\mathcal{X}}=n_x^{nr}/(N-n)$, $\tilde{H}_s=m/N_m$, $\tilde{O}_y=(n_y-n_{yx})/N_m$ and $\tilde{H}_y=\tilde{Q}_y\tilde{P}(1-\tilde{H}_s)$, where, from (4.7),

$$\begin{aligned} \tilde{Q}_y &= \frac{1}{N_m}\sum_{i=1}^{N_m}((1-s_i)i_i^{\mathcal{X}}+s_i)\left(1-(1-\tilde{H}_s-\frac{1}{1-\tilde{G}}\sum_{v\in\mathcal{Y}}\tilde{O}_v)(1-\tilde{G}^{\mathcal{X}})\right. \\ &\quad\left.-\sum_{v\in\mathcal{Y}}\tilde{O}_v\frac{\mathcal{P}\{v|x_i,\tilde{\theta}\}}{\tilde{Q}_v}\right)^{-1}\mathcal{P}\{y|x,\tilde{\theta}\},y\in\mathcal{Y}. \end{aligned}$$

and, from (4.6), the MCAR estimator $\tilde{\theta}$ satisfies[12]

$$\begin{aligned} 0 &= \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{Y}}i_i^{\mathcal{X}}\frac{\partial\log\mathcal{P}\{y_i|x_i,\tilde{\theta}\}}{\partial\theta}+((1-s_i)i^{\mathcal{X}}+s_i)\left(1-(1-\tilde{H}_s-\frac{1}{1-\tilde{G}}\sum_{v\in\mathcal{Y}}\tilde{O}_v)(1-\tilde{G}^{\mathcal{X}})\right. \\ &\quad\left.-\sum_{v\in\mathcal{Y}}\tilde{O}_v\frac{\mathcal{P}\{v|x_i,\tilde{\theta}\}}{\tilde{Q}_v}\right)^{-1}\sum_{v\in\mathcal{Y}}\tilde{O}_v\frac{1}{\tilde{Q}_v}\frac{\partial\mathcal{P}\{v|x_i,\tilde{\theta}\}}{\partial\theta}, \end{aligned}$$

Let $\tilde{\varphi}$ denote the MCAR estimator for $\varphi$. A test for data MCAR may be based on the difference of estimated GMM criteria (4.9) under null and alternative hypotheses; i.e., the statistic

$$N_m[\hat{g}(\tilde{\varphi})'\hat{\Omega}^{-1}\hat{g}(\tilde{\varphi})-\hat{g}(\hat{\varphi})'\hat{\Omega}^{-1}\hat{g}(\hat{\varphi})], \tag{5.2}$$

where $\hat{\Omega}=\sum_{i=1}^{N_m}\hat{g}(\hat{\varphi})\hat{g}(\hat{\varphi})'/N_m$. Under the MCAR null hypothesis (5.1), the statistic (5.2) will converge in distribution to a chi-square random variable with $2(C-1)$ degrees of freedom. See Newey and West (1987) for other asymptotically equivalent test statistics.

In general, random sample ML applied to the complete sample will yield consistent estimators only if data MCAR and either $G=1$, i.e., the event $\{I^{\mathcal{Y}}=1,I^{\mathcal{X}}=0\}$ is not observed, or INR$(y)$ units are ignored and the probabilities $H_y$, $G_y=1$, $y\in\mathcal{Y}$, and $G^{\mathcal{X}}$ redefined accordingly as in section 2.3; cf. (4.6). The relevant null hypothesis data MCAR now is

$$\frac{H_y}{Q_y}=P(1-H_s),y\in\mathcal{Y}. \tag{5.3}$$

The appropriate data likelihood under the alternative now imposes $G_y=1$, $y\in\mathcal{Y}$, in (3.2) and, consequently, the moment indicator and parameter vectors $g(\varphi)$ and $\varphi$ are defined with $G_y$, $y\in\mathcal{Y}$,

[12]Note that $\sum_{v\in\mathcal{Y}}(H_v/Q_v)\partial\mathcal{P}\{v|x,\theta\}/\partial\theta=(H_y/Q_y)\sum_{v\in\mathcal{Y}}\partial\mathcal{P}\{v|x,\theta\}/\partial\theta=0$, $\sum_{y\in\mathcal{Y}}(H_y/Q_y)\mathcal{P}\{y|x,\theta\}=H_y/Q_y$ and $\sum_{y\in\mathcal{Y}}H_y=P(1-H_s)$.

deleted. The estimators under (5.3) for $P$, $G^{\mathcal{X}}$, $H_s$, $O_y$ and $H_y$ are defined as above with $\tilde{\theta}$ now the random sample ML estimator on $\{I^{\mathcal{Y}}=1, I^{\mathcal{X}}=1\}$ units and $\tilde{Q}_y = \sum_{i=1}^{N_m}((1-s_i)i_i^{\mathcal{X}} + s_i)\mathcal{P}\{y|x_i,\tilde{\theta}\}/(n_m+n_x^{nr})$. Let $\hat{\varphi}$ and $\tilde{\varphi}$ denote the unrestricted and (5.3) estimators for $\varphi$. Under the null hypothesis (5.3), the statistic (5.2) converges in distribution to a chi-square random variable with $C-1$ degrees of freedom.

## 5.2 Missing Data Mechanism

The conditional independence Assumptions 2.3 and 2.4 of section 2.3 is crucial to the foregoing analysis. We initially concentrate on a useful diagnostic for the credibility of Assumption 2.3 with a specification test for Assumption 2.4 briefly outlined at the end of the section. Consider the generalisation of (2.3) given by

$$P_y(x) = \mathcal{P}\{I^{\mathcal{Y}}=1|Y=y, X=x\}.$$

The sample density (3.2) is then modified as

$$\begin{aligned} h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) \;=\; & \left\{\left[\left(\frac{P_y(x)}{P_y}\frac{H_y}{Q_y}G_y\mathcal{P}\{y|x,\theta\}f_X(x)\right)^{i^{\mathcal{X}}}\right.\right. \\ & \times\left(\frac{H_y}{Q_y}(1-G_y)\int_{\mathcal{X}}\frac{P_y(x)}{P_y}\mathcal{P}\{y|x,\theta\}f_X(x)dx\right)^{(1-i^{\mathcal{X}})}\Bigg]^{i^{\mathcal{Y}}} \\ & \times\left[\left(\sum_{v\in\mathcal{Y}}\frac{1-P_v(x)}{1-P_v}(1-H_s-\frac{H_v}{Q_v})\mathcal{P}\{v|x,\theta\}G^{\mathcal{X}}f_X(x)\right)^{i^{\mathcal{X}}}\right. \\ & \left.\left.\times\left(\sum_{v\in\mathcal{Y}}\int_{\mathcal{X}}\frac{1-P_v(x)}{1-P_v}(1-H_s-\frac{H_v}{Q_v})\mathcal{P}\{v|x,\theta\}(1-G^{\mathcal{X}})f_X(x)dx\right)^{(1-i^{\mathcal{X}})}\right]^{(1-i^{\mathcal{Y}})}\right\}^{1-s} \\ & \times[H_sf_X(x)]^s. \end{aligned}$$

The proposed specification test for Assumption 2.3 is based on the Lagrange multiplier principle; see *inter alia* Newey and West (1987). First, the probability $P_y(x)$ is parameterised as $P_y(x) = P_y(z_y'\eta_y)$ where $z_y = z_y(x)$ is a vector of independent functions of the covariates and $P_y(0) = P_y$, $y \in \mathcal{Y}$. Secondly, log-likelihoods are constructed based on the sample density (5.4); cf. (4.1). Thirdly, the moment indicator corresponding to $\eta_y$ is obtained by differentiating the resultant log-likelihood with evaluation at $\eta_y = 0$, $y \in \mathcal{Y}$; *viz.*

$$\begin{aligned} \eta_t \;:\; & P_t'(0)Q_t(1-H_s)(1-s)\left[z_ti^{\mathcal{X}}\left(i^{\mathcal{Y}}\frac{1}{H_t}\mathrm{I}(y=t)-(1-i^{\mathcal{Y}})\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}\frac{\mathcal{P}\{t|x,\theta\}}{Q_t}\right)\right. \\ & \left.+\,E[z_t|Y=t](1-i^{\mathcal{X}})\left(i^{\mathcal{Y}}\frac{1}{H_t}\mathrm{I}(y=t)-(1-i^{\mathcal{Y}})\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right)\right], \end{aligned} \tag{5.4}$$

$t \in \mathcal{Y}$, where $P_t'(\cdot)$ denotes the derivative of $P_t(\cdot)$ with respect to its argument. Note that the first term of (5.4) is proportional to $z_t$ multiplied by the component of the $H_t$ moment indicator (4.2) for

$x$-respondent units whereas the second term is $E[z_t|Y=t]$ times the component in (4.2) relevant for $x$-nonrespondent units. Fourthly, to creat an operational statistic only the first term in (5.4) suitably reweighted is retained with the observation invariant proportional factor $P_t'(0)Q_t(1-H_s)$ omitted, i.e.,

$$\eta_t : (1-s)z_t i^{\mathcal{X}} \left( i^{\mathcal{Y}} \frac{1}{G_t H_t} \mathrm{I}(y=t) - (1-i^{\mathcal{Y}}) \frac{1}{1-H_s-\sum_{v\in\mathcal{Y}} \frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}} \frac{\mathcal{P}\{t|x,\theta\}}{G^{\mathcal{X}} Q_t} \right), t \in \mathcal{Y}. \tag{5.5}$$

Note that observation of $\mathrm{INR}(x)$ sample units is required for the moment indicator (5.5). A simple modification of (5.5) appropriate for all missingness patterns is given by[13]

$$\eta_t : z_t \left( (1-s) i^{\mathcal{X}} i^{\mathcal{Y}} \frac{1}{G_t H_t} \mathrm{I}(y=t) - \frac{\mathcal{P}\{t|x,\theta\}}{Q_t} \right), t \in \mathcal{Y}. \tag{5.6}$$

Let $q(\varphi)$ define the vector of moment indicators obtained from $g(\varphi)$ defined in section 4.2 augmented by (5.5) or (5.6). Let $Q(\varphi) = \partial q(\varphi)/\partial \varphi'$. Correspondingly, define $\hat{q}(\varphi) = \sum_{i=1}^{N_m} q_i(\varphi)/N_m$, $\hat{Q}(\varphi) = \sum_{i=1}^{N_m} Q_i(\varphi)/N_m$ and $\hat{\Sigma}(\varphi) = \sum_{i=1}^{N_m} q_i(\varphi) q_i(\varphi)'/N_m$. Therefore, a GMM Lagrange multiplier specification test for Assumption 2.3 is given by

$$\mathcal{LM} = N_m \hat{q}(\hat{\varphi})' \hat{\Sigma}(\hat{\varphi})^{-1} \hat{Q}(\hat{\varphi}) \left( \hat{Q}(\hat{\varphi})' \hat{\Sigma}(\hat{\varphi})^{-1} \hat{Q}(\hat{\varphi}) \right)^{-1} \hat{Q}(\hat{\varphi}) \hat{\Sigma}(\hat{\varphi})^{-1} \hat{q}(\hat{\varphi}); \tag{5.7}$$

cf. Newey and West (1987). If Assumptions 2.3 and 2.4 are satisfied $\mathcal{LM}$ has a limiting chi-square distribution with degrees of freedom $\sum_{v\in\mathcal{Y}} \dim(\eta_v)$.

A diagnostic for Assumption 2.4 is designed similarly by generalising (2.4) and (2.5) respectively as $\mathcal{P}\{I^{\mathcal{X}}=1|I^{\mathcal{Y}}=1, Y=y, X=x\} = G_y(z_{Gy}'\eta_{Gy})$ and $\mathcal{P}\{I^{\mathcal{X}}=1|I^{\mathcal{Y}}=0, Y=y, X=x\} = G_y((z^{\mathcal{X}})'\eta^{\mathcal{X}})$ where $z_{Gy} = z_{Gy}(x)$ and $z^{\mathcal{X}} = z^{\mathcal{X}}(x)$ with $G_y(0) = G_y$, $y \in \mathcal{Y}$, $G^{\mathcal{X}}(0) = G^{\mathcal{X}}$. Operational moment indicators are then defined as $z_{Gt}(1-s)i^{\mathcal{Y}}(i^{\mathcal{X}} - G_t)\mathrm{I}(y=t)$, $t \in \mathcal{Y}$, and $z^{\mathcal{X}}(1-s)(1-i^{\mathcal{Y}})(i^{\mathcal{X}} - G^{\mathcal{X}})$, i.e., $z_t$ and $z^{\mathcal{X}}$ multiplied by the moment indicators for $G_t$ (4.3) and $G^{\mathcal{X}}$ (4.4).[14] A GMM Lagrange multiplier specification test for Assumption 2.4 is then defined as in $\mathcal{LM}$ (5.7) and has a limiting chi-square distribution with degrees of freedom $\sum_{v\in\mathcal{Y}} \dim(\eta_{Gv}) + \dim(\eta^{\mathcal{X}})$ under Assumptions 2.3 and 2.4.

# 6 Extensions

## 6.1 Unknown $N$

To adapt analysis to the unknown $N$ case, one of the nonrespondent categories should be suppressed. E.g., in the absence of UNR units, the density function of the observed data is, of course, is identical

[13] The implicit null hypotheses corresponding to (5.5) and (5.6) under Assumption 2.4 are the sub-hypotheses

$$E[z_t \left( \frac{P_t(x)}{P_t} - \frac{\sum_{v\in\mathcal{Y}}(1-P_v(x))\mathcal{P}\{v|x,\theta\}}{\sum_{v\in\mathcal{Y}}(1-P_v)\mathcal{P}\{v|x,\theta\}} \right) |Y = t] = 0,$$

$$E[z_t \left( \frac{P_t(x)}{P_t} - 1 \right) |Y = t] = 0, t \in \mathcal{Y}.$$

[14] The implicit null hypothesis corresponding to these moment indicators comprises the sub-hypotheses $E[z_{Gt}(G_t(x) - G_t)|Y=t] = 0$, $t \in \mathcal{Y}$, and $E[z^{\mathcal{X}}(G^{\mathcal{X}}(x) - G^{\mathcal{X}})] = 0$.

in form to that for pure INR; see (3.3). It is important, however, to note that the probabilities $H_y$, $G_y$ and $G^{\mathcal{X}}$ in (3.2) are now defined conditionally on the complement of $\{I^{\mathcal{Y}}=0, I^{\mathcal{X}}=0\}$; in particular, $G^{\mathcal{X}}=1$. Clearly, Assumption 2.4 relating to $G^{\mathcal{X}}$ is no longer required.

If, additionally, there are no item nonrespondents, probabilities are conditional on $\{I^{\mathcal{Y}}=1, I^{\mathcal{X}}=1\}$. Hence, $G_y=1$, $y\in\mathcal{Y}$, and Assumption 2.4 may be dropped completely. Therefore, redefining the stratum sampling probability $H_yG_y$ as $H_y=\mathcal{P}\{Y=y, S=0|I^{\mathcal{Y}}=1, I^{\mathcal{X}}=1\}$, the density function (3.3) becomes

$$h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s)=\left[\frac{H_y}{Q_y}\mathcal{P}\{y|x,\theta\}f_X(x)\right]^{i^{\mathcal{Y}}i^{\mathcal{X}}(1-s)}[H_sf_X(x)]^s. \tag{6.1}$$

Now $\mathcal{P}\{S=0|I^{\mathcal{Y}}=1, I^{\mathcal{X}}=1\}=\sum_{v\in\mathcal{Y}}H_v$, and, thus, from Assumption 2.2, $1-H_s=\sum_{v\in\mathcal{Y}}H_v$ since $\mathcal{P}\{S=0\}=\mathcal{P}\{S=0|I^{\mathcal{Y}}=1, I^{\mathcal{X}}=1\}$. Note that $Q_y$ may no longer be written in terms of $H_y$ as in (2.9). Furthermore, the relation between $P_y$, $H_y$, $H_s$ and $Q_y$, noting (2.6), is now given by

$$H_y=\frac{P_yQ_y\,(1-H_s)}{\sum_{v\in\mathcal{Y}}P_vQ_v}; \tag{6.2}$$

cf. (2.10). Therefore, $H_y$ is no longer necessarily less than $Q_y$. Moreover, even if $H_y$ and $Q_y$ are known, the probabilities $P_y$ are not identified although their ratios are from $P_{y_1}/P_{y_2}=(H_{y_1}/H_{y_2})/(Q_{y_1}/Q_{y_2})$, which is, of course, 1 all $y$ for data MCAR. In contrast with the known $N$ case, $H_y=Q_y\,(1-H_s)$ all $y$ characterizes both data MCAR *and* the absence of missing data.[15]

The density function (6.1) coincides with that for CB sampling with (without) a SRS, see, e.g., Cosslett (1981a), and in the absence of a SRS corresponds to exactly that examined in Tang *et al.* (2003) when adapted for our discrete outcome setting. Inference procedures appropriate for CB samples may therefore be used when $N$ is unknown and in the absence of item nonrespondents. Hence, our estimator, when simplified to deal with this case, coincides with Imbens' (1992). Similarly, Cosslett's (1981a) ML estimator for CB samples combined with a SRS of covariates, may be employed to describe unit nonresponse if information on $N$ is ignored. However, in the same sense that Imbens (1992) simplified Cosslett's (1981a, b) estimator for CB samples, the GMM estimator for UNR derived here is substantially simpler than that corresponding to Cosslett (1981a).

Furthermore, our estimator embeds Lancaster and Imbens' (1996) efficient GMM estimator for case-control binary models with contaminated controls, where there are two strata, one consisting of a random sample where only the covariates are observable, the other including units choosing $Y=2$ where $X$ is fully observed.[16]

[15]The sample density under MCAR now becomes

$$h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s)=\mathcal{P}\{y|x,\theta\}^{i^{\mathcal{Y}}i^{\mathcal{X}}(1-s)}\left[H_s^s(1-H_s)^{i^{\mathcal{Y}}i^{\mathcal{X}}(1-s)}\right]f_X(x)^{i^{\mathcal{Y}}i^{\mathcal{X}}(1-s)+s},$$

with the MCAR estimators $\tilde{\theta}$ the random sample ML estimator, $\tilde{H}_s=m/n_m$ and, from (4.7), $\tilde{Q}_y=\sum_{i=1}^{n_m}\mathcal{P}\{y|x_i,\tilde{\theta}\}/n_m$, $y\in\mathcal{Y}$. Under the MCAR null hypothesis $H_0: H_y=Q_y\,(1-H_s)$, $y\in\mathcal{Y}$, the statistic (5.2) will converge in distribution to a chi-square random variable with $C$ degrees of freedom.

[16]In this case, $\mathcal{Y}=\{1,2\}$, $P_1=0$ and $P_2=1$. Hence, Assumption 2.3 where $0<P_y<1$ is relaxed to $0\leq P_y\leq 1$.

## 6.2 Information on $Q_y$

To incorporate additional information on the marginal probabilities $Q_y$, $y \in \mathcal{Y}$, we follow the approach of Imbens and Lancaster (1994). Such information may arise, e.g., from an independent survey or a census.

Let $\hat{Q} = (\hat{Q}_1, ..., \hat{Q}_C)'$ be an estimator for the marginal choice probabilities $Q = (Q_1, ..., Q_C)'$ based on a sample independent from the initial and supplementary random samples. Partition $\varphi$ as $\varphi = (\phi', Q')'$.

**Assumption 6.1** *(Additional Information on Marginal Choice Probabilities.) The estimator $\hat{Q}$ for the marginal choice probabilities $Q^0$ is independent of the initial and supplementary random samples and satisfies*

$$N_Q^{1/2}(\hat{Q} - Q^0) \xrightarrow{d} N(0, \Sigma^0),$$

*where $N_Q$ is known, $N_m/N_Q \to \rho_Q$, $0 < \rho_Q < \infty$, and $\Sigma$ is nonsingular. A consistent estimator $\hat{\Sigma}$ of $\Sigma^0$ is assumed to be available.*

Let $\hat{h}(\varphi) = (\hat{g}(\varphi)', (\hat{Q} - Q)')'$. Consider the GMM criterion

$$\tilde{J}(\varphi) = \hat{h}(\varphi)' \hat{\Xi}^{-1} \hat{h}(\varphi), \tag{6.3}$$

where $\hat{\Xi} = \texttt{diag}(\hat{\Omega}, (N_m/N_Q)\hat{\Sigma})$ with $\hat{\Omega}$ a consistent estimator for $\Omega^0$, e.g., $\hat{\Omega} = \sum_{i=1}^{N_m} g_i(\dot{\varphi}) g_i(\dot{\varphi})'/N_m$ and $\dot{\varphi}$ a consistent estimator for $\varphi^0$, obtained, for example, as in section 4.2. Let $\tilde{\varphi} = (\tilde{\phi}', \tilde{Q}')'$ denote the minimiser of (6.3) and define

$$H = \begin{pmatrix} G_\phi & G_Q \\ 0 & -I_C \end{pmatrix}, \Xi = \texttt{diag}(\Omega, \rho_Q \Sigma),$$

where $G = (G_\phi, G_Q)$ is partitioned conformably with $\varphi = (\phi', Q')'$.

**Theorem 6.1** *(Consistency and Asymptotic Normality of $\tilde{\varphi}$.) If Assumptions 2.1-2.5, 4.1-4.3 and 6.3 are satisfied, then*

$$\tilde{\varphi} \xrightarrow{p} \varphi^0, N_m^{1/2} \left(\tilde{\varphi} - \varphi^0\right) \xrightarrow{d} N(0, ((H^0)'(\Xi^0)^{-1} H^0)^{-1}).$$

Asymptotic efficiency of $\tilde{\varphi}$ may be proved in a similar manner to that employed in Appendix B for the semiparametric efficiency of $\hat{\varphi}$.

If components of $\phi$ are of primary concern rather than $Q$, an asymptotically equivalent estimator to $\tilde{\phi}$ is obtained by minimisation of the following GMM objective function

$$\hat{g}(\phi, \hat{Q})'(\hat{\Omega}^{-1} - \hat{\Omega}^{-1} \hat{G}_Q (\hat{H}_Q' \hat{\Xi}^{-1} \hat{H}_Q)^{-1} \hat{G}_Q' \hat{\Omega}^{-1}) \hat{g}(\phi, \hat{Q}),$$

where $\hat{H}_Q = (\hat{G}_Q', -I_C)'$ denotes a consistent estimator for $H_Q^0 = ((G_Q^0)', -I_C)'$. Cf. Imbens and Lancaster (1994, pp.665-6).

If $\rho_Q = \infty$, then, in an asymptotic sense, $\hat{Q}$ contributes no additional information to the estimation of $\varphi^0$ over and above that in the main and supplementary samples and consequently

may be ignored. Whereas, if $\rho_Q = 0$, then $Q^0$ may be treated as known and, thus, $\hat{Q}$ may be substituted for $Q$ in system (4.2)-(4.8) rendering it over-identified. An optimal GMM estimator is then obtained using the weighting matrix $\hat{W} = \hat{\Omega}^{-1}$ in (4.9). Similarly to Theorem 6.1 and, in particular, (4.10),

$$\tilde{\phi} \xrightarrow{p} \phi^0, N_m^{1/2}(\tilde{\phi} - \phi^0) \xrightarrow{d} N(0, ((G_\phi^0)'(\Omega^0)^{-1}G_\phi^0)^{-1}).$$

# 7 Simulation Evidence

This section presents a simulation study based on an ordered Probit model to assess the performance of some of the estimators developed in previous sections. To provide a realistic setting which is likely to reflect nonresponse as it occurs in practice, we base our investigation on a dataset where the problem of nonignorable INR is well-documented; section 7.1 details the empirical basis for our experiments. Section 7.2 describes the experimental design for several alternative patterns of nonresponse based on these results and sub-samples of the dataset and section 7.3 discusses the results.

## 7.1 Empirical Results for the UK Labour Force Survey

We consider an application where nonignorable nonresponse has been considered to be a potentially serious problem. Papers such as Skinner *et al.* (2002) and Durrant and Skinner (2006) are especially interested in estimating the lower end of the distribution of hourly pay in the UK using the Labour Force Survey (LFS). Skinner *et al.* (2002) shows that when hourly pay is measured indirectly, the *derived variable*, as a ratio of weekly earnings and weekly hours worked, it may be subject to a large amount of measurement error. If hourly pay, the *direct variable*, is measured by asking the hourly paid workers their hourly rate of pay, measurement error is, in principle, eliminated or at least reduced. Measurements of the hourly rate are effectively missing for about three quarters of the sampling units, which are typically the more highly paid employees since hourly paid workers tend to have lower wages. Skinner *et al.* (2002) propose an imputation procedure to deal with the missing values of the direct variable denoted here by *hrrate*. They consider the complete sample from LFS data for the 22+ age group for June-August 1999 and, under the assumption of data missing at random (MAR), see fn. 6, use a regression of the logarithm of *hrrate* on the logarithm of the derived variable and a set of covariates to estimate the missing values of *hrrate*. They concluded that the proportion of the UK population below and at the National Minimum Wage, fixed at £3.60 in 1999 for that age group, was 5.5%.

Nonresponse here depends on the outcome variable of hourly wages. Therefore, the response mechanism described in section 2.3 may be relevant as a description of this pattern of missingness. However, our estimators are designed to deal with discrete response. Since the dependent variable typically takes only a limited number of values, we restricted consideration to 2, 3, and 4 alternatives and discretized the logarithm of *hrrate* as described in Table 1. The definition of the classes was made in such a way that the first class, where $Y = 1$, includes individuals with an hourly rate no higher than the National Minimum Wage of £3.60. Thus, the estimate of $Q_1$ provides the

proportion of the population below or at the National Minimum Wage.

**Table 1 about here**

We consider ordered probit models characterized by $\mathcal{P}\{Y = 1|x\} = \Phi(\alpha_1 - x'\theta)$, $\mathcal{P}\{Y = j|x\} = \Phi(\alpha_j - x'\theta) - \Phi(\alpha_{j-1} - x'\theta)$, $(j = 2, ..., C-1)$, and $\mathcal{P}\{Y = C|x\} = 1 - \Phi(\alpha_{C-1} - x'\theta)$, where $\alpha_j$, $(j = 1, ..., C-1)$, are the $C-1$ class limits and $\Phi(\cdot)$ denotes the standard normal distribution function. Since the class limits $\alpha_j$, $(j = 1, ..., C-1)$, are known by design, their known values are incorporated directly in the moment conditions; see, e.g., Stewart (1983).

Essentially, the covariates are the same as those of Skinner *et al.* (2002) except their *derived variable* is excluded to avoid their assumption of data MAR and we considered more aggregated occupation categories. Because nonresponse only concerns the outcome variable $Y$, we only consider pure INR($x$) nonresponse in the absence of a SRS. Hence, $G_y = 1$, $y \in \mathcal{Y}$, $G^{\mathcal{X}} = 1$ and $H_s = 0$; see section 3.1. Both random sample (RS) ML and (INR) GMM estimators are examined with the results presented in Table 2. The majority of coefficient estimates are significant at the 0.01 level. Although their magnitude is often very different between RS ML and INR GMM, their signs, with the exception of that of the parameter associated with the clerical and secretarial occupation category, are the same, coinciding with those of Table 6 in Skinner *et al.* (2002). Although it is not our purpose to propose an alternative model for these data to Skinner *et al.* (2002), it is reassuring to note that we obtain a not too dissimilar estimate of 7.7% for the fraction of the population below or at the National Minimum Wage.

**Table 2 about here**

## 7.2 Experimental Design

Initial sample sizes $N = 500$ or 1000 were considered. These sample sizes are much smaller than those typically encountered in microeconometrics; e.g., the LFS data set has approximately 16000 employees in the 22+ age group. Each experiment comprised 1000 replications, each of which were collected as a random sample with replacement from the LFS data set. The discrete outcome variable $Y$ was generated by an ordered probit model with the classes defined in Table 1 for $C = \{2, 3, 4\}$. For computational reasons, we considered only the three most significant covariates in Table 2 for INR GMM with $C = 4$ and thus set $\theta^0 = 2.293, 0.027, -0.671, 1.159)$. Table 3 details several patterns of INR, characterized by different combinations of conditional response probabilities for $Y$. To mimic the missing data pattern of the LFS values of *hrrate*, experiments were designed where it was especially missing for well-paid employees. The rate of response in Design $c$ in the higher hourly rate classes is 50%, which yields an overall rate of response, $\mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\}$, of around 54%. Design $d$ increases those rates of response to, respectively, 70% and 73%, while Design $e$ permits differential rates of response in all classes of $Y$ but in such a way that the overall rate of response is again around 54%. Designs $a$ and $b$ allow the performance of RS ML to be evaluated when no data is missing and data is MCAR for comparison with the misspecified scenarios in Designs $c$, $d$, and $e$.

**Table 3 about here**

RS ML and INR GMM estimators were computed for Designs $c$, $d$, and $e$. We also considered UNR and INR and UNR GMM estimation with known $Q_y$ for binary models. We only present partial results for known $Q_y$, since this situation is less likely to occur in practice. Indeed, the main aim of Skinner *et al.* (2002) and Durrant and Skinner (2006) is precisely the estimation of the proportion $Q_1$ of employees below or at the National Minimum Wage in the UK. With regard to UNR GMM estimation, in some examples for $C = 3$ and 4, not reported here, the performance of UNR GMM estimation was poor, displaying wide dispersion and on occasion a failure to converge, which lead us to suspect of the possibility of a lack of identification in the samples considered.[17] Such problems are circumvented by INR GMM estimation, since covariate information for nonrespondents is now available. RS ML estimators only are computed for Designs $a$ and $b$. Design $a$ should act as an especially interesting reference point in terms of precision since, as all the data is available, standard errors should be expected to assume their minima among all the designs considered. Design $b$ should illustrate the consistency of RS ML under data MCAR and provide some guide to the decay in the precision due to the effective reduction in sample size by 50%. All computations were done using `S − Plus`. Additional simulation results for binary models where the performance of RS ML, INR and UNR GMM estimators with $Q_y$ unknown and known is compared are presented in Ramalho and Smith (2003).

### 7.3 Results

Summary statistics are presented in Tables 4-6, which provide estimator proportionate mean and median bias, standard deviation and mean absolute error.

**Tables 4-6 about here**

As expected, RS ML estimation performs well in Designs $a$ and $b$ although, in the MCAR Design $b$, dispersion increases substantially due to the effective reduction in sample size of 50%. In all other designs, these estimators suffer from both large mean and median biases. These biases typically increase with the variation in response probabilities $P_y$; cf. Designs $c$ and $d$ in Tables 4-6 and Designs $d$ and $e$ in Tables 5 and 6. Biases decline somewhat when the initial sample size is increased from $N = 500$ to 1000.

In general, the performance of INR GMM estimation is excellent. In most experiments they appear both mean and median unbiased and are more precise than RS ML. In fact, the dispersion of INR GMM estimators appears very similar to that of RS ML in Design $a$, which contains no missing values. Hence, in the absence of a substantial fraction of responses, INR GMM estimation uses the available information in such a way that the attendant negative consequences in terms of

[17] Note that the UNR estimator is similar to that proposed by Imbens (1992) for CB sampling. In the CB sampling context, Cosslett (1993, pp.10-11) suggests that intercept terms in multinomial Probit models may be poorly determined even if formally identified by analogy with a similar issue that arises in multinomial logit models; see, e.g., Manski and Lerman (1977) and Imbens (1992).

precision seem almost irrelevant.[18] Also note the improvement in bias when the number of classes is increased from $C = 2$ to 3, although results are similar for $C = 3$ and 4. Naturally, for the smaller sample size of $N = 500$, results are somewhat worse both in terms of bias and dispersion. This deterioration is primarily restricted to INR GMM estimation of $\theta_1$ and to a lesser extent $\theta_3$. The mean bias for $\theta_1$ is at a maximum of 8.8% for Design $c$ in Table 4 for $N = 500$ although median bias is only 4.3%. Even so more than 50% of the bias of the RS ML estimator is removed. The large standard errors for $\theta_3$ in Designs $c$ and $d$ of Table 4 are the only cases where this statistic is substantially larger than that of RS ML. This outturn is due to only three and one replications, respectively, where $\hat{\theta}_3$ is larger than 6.0. If these replications were ignored the standard errors would reduce to 0.276 and 0.238 respectively.

Sample size is more important for UNR GMM estimation. In Table 4 with $N = 1000$, these estimators are approximately unbiased. Moreover, these results, with the exception of the intercept, indicate a nice feature of UNR GMM being less dispersed than RS ML. However, for $N = 500$, UNR GMM displays large mean biases for $\theta_0$ and $\theta_3$ together with large standard errors, especially for $\theta_3$, although median biases are much smaller with a maximum of 4.9%. Therefore, in the absence of nonrespondent covariate information, UNR GMM estimation appears reliable with sample sizes of at least $N = 1000$. For smaller sample sizes, although they display significant variability, UNR GMM is clearly superior to uncorrected RS ML estimation. However, INR GMM is clearly superior to UNR GMM estimation, demonstrating the importance of the inclusion of available nonrespondent covariate information.

Additional results, not reported here, for binary models with aggregate information on $Q_1$ suggest that INR and UNR GMM estimators behave very similarly to INR GMM in the absence of information on $Q_1$ in Table 4. Aggregate information on $Q_y$ is therefore particularly important for UNR GMM, producing gains in terms of bias and dispersion, particularly for the intercept.

**Figures 2 and 3 about here**

Figures 2 and 3 present estimated sampling densities of RS ML and INR GMM estimators for some of the above designs. Figure 2 examines RS ML and INR GMM when $C = 4$ for (I) Design $b$ and (II) Design $c$, whereas Figure 3 presents results for Design $e$ for (I) $N = 1000$ and $C = 2, 3$ and 4 and (II) $N = \{500, 1000, 2000\}$ and $C = 4$. Figure 2 confirms the above conclusions of the superiority of INR GMM over RS ML in circumstances when nonresponse is nonignorable. Moreover, for data MCAR, INR GMM does not display a noticable deficiency as compared with RS ML. Figure 3 underlines the improvement in performance for INR GMM as the number of classes $C$ or initial sample size $N$ increase.

These experiments and those in Ramalho and Smith (2003) confirm that, in general, RS ML estimation on the complete sample is only sensible when nonresponse is ignorable. Otherwise, these

[18] Estimators that correct for sampling issues like missing data or measurement error are generally expected to be more disperse than uncorrected biased estimators, reflecting the additional variability of the data. For INR GMM estimators this loss of precision is circumvented by the inclusion of information on covariates provided by nonrespondents, which is not used in RS ML estimation.

estimators are inappropriate. The performance of both INR and UNR GMM estimators is very promising. Aggregate information on $Q_y$ is especially beneficial for UNR GMM.

# 8 Conclusions

This paper considers a general framework for missing data when the dependent variable is discrete. A unified semiparametrically efficient GMM estimation and inference methodology is proposed for such circumstances which adapts and extends that usually employed with choice based sampling. The advantages of an integrated approach are clear with the same methodology being employed for both model specification and estimator derivation in all cases. Additionally, the investigation of and comparison between different nonresponse patterns and problems is straightforward, e.g., specialisation to pure unit or item nonresponse.

The critical assumption in our framework, besides the correct specification of the structural model, concerns the independence of response and covariates conditional on the discrete outcome variable. This assumption might be expected to be relevant in a number of practical situations. In INR($y$), it is not necessarily too unreasonable to assume that covariates influence the choice variable and the willingness to report that choice in a similar fashion. For UNR, this assumption is likely to be appropriate in cases where the refusal to participate in the survey is especially motivated by an unwillingness to reveal the value of the choice variable. We suggest how this assumption may be weakened to allow the response mechanism to depend additionally in a discrete fashion on covariates. Specification tests are presented both for MCAR and the missingness assumption.

A simulation study reveals very promising results. The GMM estimators suggested here display negligible bias, which is especially apparent in cases where data on the covariates from nonrespondents are incorporated in the estimation procedure. In contradistinction, random sample ML estimators are considerably biased in all cases where response rates across the alternative choices are different, even in experiments where the differential was not very substantial. The incorporation of aggregate information on the marginal population choice probabilities is especially advantageous for the properties of the unit nonresponse GMM estimator.

# Appendix A: Derivation of Moment Indicators

Let $\mathcal{L}$ denote the Lagrangean arising from (4.1) with $\mu$ the Lagrange multiplier associated with the constraint $\sum_{x\in\mathcal{X}} \pi_x = 1$. The resultant first order derivatives are

$$\frac{\partial \mathcal{L}}{\partial H_y} = \sum_{i=1}^{N_m}(1-s_i)\left[i_i^{\mathcal{Y}}\frac{\mathrm{I}(y_i=y)}{H_y}\right. \tag{A.1}$$
$$-(1-i_i^{\mathcal{Y}})\left(i_i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}}\frac{\mathcal{P}\{y|x_i,\theta\}}{Q_y}\right.$$
$$\left.\left.+(1-i_i^{\mathcal{X}})\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right)\right],$$

$$\frac{\partial \mathcal{L}}{\partial G_y} = \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{Y}}\mathrm{I}(y_i=y)\left[i_i^{\mathcal{X}}\frac{1}{G_y}-(1-i_i^{\mathcal{X}})\frac{1}{1-G_y}\right], \tag{A.2}$$

$$\frac{\partial \mathcal{L}}{\partial G^{\mathcal{X}}} = \sum_{i=1}^{N_m}(1-s_i)(1-i_i^{\mathcal{Y}})\left[i_i^{\mathcal{X}}\frac{1}{G^{\mathcal{X}}}-(1-i_i^{\mathcal{X}})\frac{1}{1-G^{\mathcal{X}}}\right], \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial H_s} = \sum_{i=1}^{N_m}s_i\frac{1}{H_s}-(1-s_i)(1-i_i^{\mathcal{Y}})\left[i_i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}}\right. \tag{A.4}$$
$$\left.+\left(1-i_i^{\mathcal{X}}\right)\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\right],$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{X}}\left[i_i^{\mathcal{Y}}\left(\frac{\partial\log\mathcal{P}\{y_i|x_i,\theta\}}{\partial\theta}-\frac{1}{Q_{y_i}}\sum_{x\in\mathcal{X}}\pi_x\frac{\partial\mathcal{P}\{y_i|x,\theta\}}{\partial\theta}\right)\right. \tag{A.5}$$
$$-(1-i_i^{\mathcal{Y}})\left(\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}}\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\frac{\partial\mathcal{P}\{v|x_i,\theta\}}{\partial\theta}\right.$$
$$\left.\left.-\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}}\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v^2}\mathcal{P}\{v|x_i,\theta\}\sum_{x\in\mathcal{X}}\pi_x\frac{\partial\mathcal{P}\{v|x,\theta\}}{\partial\theta}\right)\right],$$

$$\frac{\partial \mathcal{L}}{\partial \pi_x} = \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{Y}}\left[i_i^{\mathcal{X}}\left(\frac{\mathrm{I}(x_i=x)}{\pi_x}-\frac{1}{Q_{y_i}}\mathcal{P}\{y_i|x,\theta\}\right)\right. \tag{A.6}$$
$$+(1-s_i)(1-i_i^{\mathcal{Y}})i_i^{\mathcal{X}}\left(\frac{\mathrm{I}(x_i=x)}{\pi_x}+\frac{\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v^2}\mathcal{P}\{v|x_i,\theta\}\mathcal{P}\{v|x,\theta\}}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x_i,\theta\}}\right)\Bigg]$$
$$+s_i\frac{\mathrm{I}(x_i=x)}{\pi_x}-\mu,$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{x\in\mathcal{X}}\pi_x-1. \tag{A.7}$$

Equating (A.1) to zero, i.e. $\partial\mathcal{L}/\partial H_y=0$, the ML estimator for $H_y$ solves

$$\frac{n_y}{\hat{H}_y} = n_u\frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v} \tag{A.8}$$
$$+\sum_{i=1}^{N_m}(1-s_i)(1-i_i^{\mathcal{Y}})i_i^{\mathcal{X}}\frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x_i,\hat{\theta}\}}\frac{\mathcal{P}\{y|x_i,\hat{\theta}\}}{\hat{Q}_y}.$$

Likewise, from (A.2) and (A.3), i.e., $\partial\mathcal{L}/\partial G_y = 0$ and $\partial\mathcal{L}/\partial G^{\mathcal{X}} = 0$, $\hat{G}_y = n_{yx}/n_y$ and $\hat{G}^{\mathcal{X}} = n_x^{nr}/(N-n)$. The statistic $\hat{H}_s = m/N_m$ is the ML estimator for $H_s$ and is obtained by first multiplying (A.1) by $H_y$, summing over $y$ and equating the resultant expression to zero which yields

$$N - n_u \frac{1-\hat{H}_s}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v} = (1-\hat{H}_s)\sum_{i=1}^{N_m}(1-s_i)\left(1-i_i^{\mathcal{Y}}\right)i_i^{\mathcal{X}}\frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x_i,\hat{\theta}\}}$$

and then, secondly, equating (A.4) to zero.

The mass point probabilities $\pi_x$, $x \in \mathcal{X}$, can be concentrated out, thus removing the dependence on the discrete distribution of $X$; cf. Imbens (1992). Multiplying $\partial\mathcal{L}/\partial\pi_x$ (A.6) through by $\pi_x$, summing over the points of support $x \in \mathcal{X}$, equating to zero and using (A.8) yields

$$\hat{\mu} = N_m - n_u \frac{1-\hat{H}_s}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}.$$

Substituting back for $\hat{\mu}$ in (A.6) and again using (A.8)

$$\begin{aligned}\hat{\pi}_x &= \sum_{i=1}^{N_m}((1-s_i)i_i^{\mathcal{X}}+s_i)\mathrm{I}(x_i=x)\left(N_m - n_u\frac{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x,\hat{\theta}\}}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}\right.\\ &\quad\left.-\sum_{v\in\mathcal{Y}}(n_v-n_{vx})\frac{\mathcal{P}\{v|x,\hat{\theta}\}}{\hat{Q}_v}\right)^{-1}.\end{aligned}$$

Hence, the ML estimator for $Q_y$ is given by

$$\begin{aligned}\hat{Q}_y &= \sum_x \hat{\pi}_x\mathcal{P}\{y|x,\hat{\theta}\} \qquad (A.9)\\ &= \sum_{i=1}^{N_m}((1-s_i)i_i^{\mathcal{X}}+s_i)\left(N_m - n_u\frac{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x,\hat{\theta}\}}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}\right.\\ &\quad\left.-\sum_{v\in\mathcal{Y}}(n_v-n_{vx})\frac{\mathcal{P}\{v|x_i,\hat{\theta}\}}{\hat{Q}_v}\right)^{-1}\mathcal{P}\{y|x_i,\hat{\theta}\},\end{aligned}$$

$y \in \mathcal{Y}$.

Similarly, after substituting for $\hat{\pi}_x$, $x \in \mathcal{X}$, in (A.5),

$$\begin{aligned}0 &= \sum_{i=1}^{N_m}(1-s_i)i_i^{\mathcal{Y}}i_i^{\mathcal{X}}\frac{\partial\log\mathcal{P}\{y_i|x_i,\hat{\theta}\}}{\partial\theta} \qquad (A.10)\\ &\quad+((1-s_i)i_i^{\mathcal{X}}+s_i)\\ &\quad\times\left(N_m - n_u\frac{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x_i,\hat{\theta}\}}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}-\sum_{v\in\mathcal{Y}}(n_v-n_{vx})\frac{\mathcal{P}\{v|x_i,\hat{\theta}\}}{\hat{Q}_v}\right)^{-1}\end{aligned}$$

$$\times \sum_{v\in\mathcal{Y}} \left( n_v - n_{vx} - n_u \frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}\hat{H}_v \right) \frac{1}{\hat{Q}_v}\frac{\partial \mathcal{P}\{v|x_i,\hat{\theta}\}}{\partial\theta}$$

$$- (1-s_i)(1-i_i^{\mathcal{Y}})i_i^{\mathcal{X}} \frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\mathcal{P}\{v|x_i,\hat{\theta}\}} \sum_{v\in\mathcal{Y}}\frac{\hat{H}_v}{\hat{Q}_v}\frac{\partial \mathcal{P}\{v|x_i,\hat{\theta}\}}{\partial\theta} \Bigg].$$

## Appendix B: Semiparametric Efficiency

Following Imbens (1992), we construct a sequence of parametric models which satisfy the same regularity conditions as our model that always includes the semiparametric model. Estimator efficiency can be proved by showing that the Cramér-Rao lower bound associated with this model sequence converges to the asymptotic covariance matrix of our semiparametric estimator. For simplicity of exposition we assume there is no auxiliary sample information on the density $f_X(\cdot)$, i.e., $S = 0$ and $H_s = 0$.

To construct the sequence of parametric models recall that $X$ has density $f_X(\cdot)$ defined on $\mathcal{X}$. For any $\varepsilon > 0$, partition $\mathcal{X}$ into $L_\varepsilon$ subsets $\mathcal{X}_l$, $l = 1, ..., L_\varepsilon$, where $\mathcal{X}_l \cap \mathcal{X}_m = \emptyset$ if $l \neq m$ and $\|x - z\| < \varepsilon$ if $x, z \in \mathcal{X}_l$. Define $\phi_l(x) = 1$ if $x \in \mathcal{X}_l$ and 0 otherwise and $f_X^\varepsilon(x) = f_X(x)/\left[\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l} f_X(x)dx\right]$. Define the parameters $\delta_l = \mathcal{P}\{x \in \mathcal{X}_l\} = \int_{\mathcal{X}_l} f_X(x)dx$, $l = 1, ..., L_\varepsilon$.

The sequence of parametric models indexed by $\varepsilon$ is given by

$$\begin{aligned} h^\varepsilon(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s) \;=\; & \left\{ \left[ \left( H_y G_y \frac{\mathcal{P}\{y|x,\theta\} f_X^\varepsilon(x)\sum_{l=1}^{L_\varepsilon}\delta_l\phi_l(x)}{\sum_{l=1}^{L_\varepsilon}\delta_l\int_{\mathcal{X}_l}\mathcal{P}\{y|x,\theta\}f_X^\varepsilon(x)dx} \right)^{i^{\mathcal{X}}} (H_y(1-G_y))^{(1-i^{\mathcal{X}})} \right]^{i^{\mathcal{Y}}} \right. \\ & \times \left[ \left( (1-H_s-\sum_{v\in\mathcal{Y}} H_v \frac{\mathcal{P}\{v|x,\theta\}}{\sum_{l=1}^{L_\varepsilon}\delta_l\int_{\mathcal{X}_l}\mathcal{P}\{v|x,\theta\}f_X^\varepsilon(x)dx}) G^{\mathcal{X}} f_X^\varepsilon(x) \sum_{l=1}^{L_\varepsilon}\delta_l\phi_l(x) \right)^{i^{\mathcal{X}}} \right. \\ & \left. \left. \times \left( (1-H_s-\sum_{v\in\mathcal{Y}} H_v)(1-G^{\mathcal{X}}) \right)^{(1-i^{\mathcal{X}})} \right]^{(1-i^{\mathcal{Y}})} \right\}^{1-s} \left[ H_s f_X^\varepsilon(x) \sum_{l=1}^{L_\varepsilon}\phi_l(x)\delta_l \right]^s, \end{aligned}$$

with $f_X^\varepsilon(x)$ a known function and $H_y$, $G_y$, $y \in \mathcal{Y}$, $G^{\mathcal{X}}$, $\theta$ and $\delta_l$, $l = 1, ..., L_\varepsilon$, the unknown parameters.

The ML estimator for $Q_y$ from (2.2) is defined as

$$\hat{Q}_y = \sum_{l=1}^{L_\varepsilon}\hat{\delta}_l \int_{\mathcal{X}_l}\mathcal{P}\{y|x,\hat{\theta}\}f_X^\varepsilon(x)dx.$$

Hence, the dependence of the likelihood equations obtained from $h^\varepsilon(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},s)$ on $\delta_l$ may be removed by the same procedure employed to remove dependence on $\hat{\pi}_x$ in the system (A.1)-(A.7).

The resultant score vector is described by the moment indicators

$$
\begin{aligned}
H_t \;:\; & (1-s)i^{\mathcal{Y}}\frac{1}{H_t}\mathrm{I}(y=t)-(1-s)(1-i^{\mathcal{Y}})(1-i^{\mathcal{X}})\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}H_v} \qquad \text{(B.1)}\\
& -(1-s)(1-i^{\mathcal{Y}})i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}\frac{\mathcal{P}\{t|x,\theta\}}{Q_t},\ t\in\mathcal{Y},\\
G_t \;:\; & (1-s)i^{\mathcal{Y}}(i^{\mathcal{X}}-G_t)\mathrm{I}(y=t),\ t\in\mathcal{Y}, \qquad \text{(B.2)}\\
G^{\mathcal{X}} \;:\; & (1-s)(1-i^{\mathcal{Y}})(i^{\mathcal{X}}-G^{\mathcal{X}}), \qquad \text{(B.3)}\\
H_s \;:\; & s-H_s, \qquad \text{(B.4)}\\
\theta \;:\; & (1-s)i^{\mathcal{X}}i^{\mathcal{Y}}\frac{\partial\log\mathcal{P}\{y|x,\theta\}}{\partial\theta} \qquad \text{(B.5)}\\
& +((1-s)i^{\mathcal{X}}+s)\left(1-(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_y})(1-G^{\mathcal{X}})\right.\\
& \times\frac{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\mathcal{P}\{v|x,\theta\}f_X^\varepsilon(x)dx}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\\
& \left.-\sum_{v\in\mathcal{Y}}O_y\frac{1}{Q_v}\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\mathcal{P}\{v|x,\theta\}f_X^\varepsilon(x)dx\right)^{-1}\sum_{v\in\mathcal{Y}}\left(O_y-\frac{1}{1-\hat{H}_s-\sum_{v\in\mathcal{Y}}\hat{H}_v}\right.\\
& \left.\times\,(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_y})(1-G^{\mathcal{X}})H_v\right)\frac{1}{Q_v}\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\frac{\partial\mathcal{P}\{v|x,\theta\}}{\partial\theta}f_X^\varepsilon(x)dx\\
& \times\,-(1-s)(1-i^{\mathcal{Y}})i^{\mathcal{X}}\frac{1}{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\mathcal{P}\{v|x,\theta\}}\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\frac{\partial\mathcal{P}\{v|x,\theta\}}{\partial\theta},
\end{aligned}
$$

$$
\begin{aligned}
Q_y \;:\; & Q_y \qquad \text{(B.6)}\\
& -((1-s)i^{\mathcal{X}}+s)\left(1-(1-H_s-\sum_{v\in\mathcal{Y}}\frac{O_v}{1-G_y})(1-G^{\mathcal{X}})\right.\\
& \times\frac{1-H_s-\sum_{v\in\mathcal{Y}}\frac{H_v}{Q_v}\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\mathcal{P}\{v|x,\theta\}f_X^\varepsilon(x)dx}{1-H_s-\sum_{v\in\mathcal{Y}}H_v}\\
& \left.-\sum_{v\in\mathcal{Y}}O_y\frac{1}{Q_v}\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\mathcal{P}\{v|x,\theta\}f_X^\varepsilon(x)dx\right)^{-1}\\
& \times\sum_{l=1}^{L_\varepsilon}\phi_l(x_i)\int_{\mathcal{X}_l}\mathcal{P}\{y|x,\theta\}f_X^\varepsilon(x)dx,
\end{aligned}
$$

$$
O_t : (1-s)i^{\mathcal{Y}}(1-i^{\mathcal{X}})\mathrm{I}(y=t)-O_t. \qquad \text{(B.7)}
$$

Define the expectation $\mathcal{E}_\varepsilon[\mathcal{P}\{y|x,\theta\}]=\sum_{l=1}^{L_\varepsilon}\phi_l(x)\int_{\mathcal{X}_l}\mathcal{P}\{y|x,\theta\}f_X^\varepsilon(x)dx$ and $\mathcal{E}_\varepsilon[\partial\mathcal{P}\{y|x,\theta\}/\partial\theta]$, $\mathcal{E}_\varepsilon[\partial^2\mathcal{P}\{y|x,\theta\}/\partial\theta\partial\theta']$ similarly. The difference between the moment indicators (B.1)-(B.7) and (4.2)-(4.8) is that the respective expectations replace $\mathcal{P}\{y|x,\theta\}$ and $\partial\mathcal{P}\{y|x,\theta\}/\partial\theta$.

Continuous differentiability of $\mathcal{P}\{y|x,\theta\}$, $\partial\mathcal{P}\{y|x,\theta\}/\partial\theta$ and $\partial^2\mathcal{P}\{y|x,\theta\}/\partial\theta\partial\theta'$ in $x$ implies uniform convergence of $\mathcal{E}_\varepsilon[\mathcal{P}\{y|x,\theta\}]$, $\mathcal{E}_\varepsilon[\partial\mathcal{P}\{y|x,\theta\}/\partial\theta]$ and $\mathcal{E}_\varepsilon[\partial^2\mathcal{P}\{y|x,\theta\}/\partial\theta\partial\theta']$ to $\mathcal{P}\{y|x,\theta\}$, $\partial\mathcal{P}\{y|x,\theta\}/\partial\theta$ and $\partial^2\mathcal{P}\{y|x,\theta\}/\partial\theta\partial\theta'$ respectively. Let $\Omega_\varepsilon = E[g^\varepsilon(\varphi^0)g^\varepsilon(\varphi^0)']$ and $G_\varepsilon = E[\partial g^\varepsilon(\varphi^0)/\partial\varphi']$ where $g^\varepsilon(\varphi)$ stacks the moment indicators (B.1)-(B.7). Hence, $\lim_{\varepsilon\to 0}\Omega_\varepsilon = \Omega$ and $\lim_{\varepsilon\to 0}G_\varepsilon = G$. Thus, the asymptotic variance matrix $G_\varepsilon^{-1}\Omega_\varepsilon G_\varepsilon'^{-1}$, which is the Cramér-Rao lower bound for the parametric ML estimator defined by (B.1)-(B.7), also converges to $G^{-1}\Omega G'^{-1}$, the asymptotic variance matrix of the GMM estimator. Therefore, the GMM estimator is semiparametrically efficient.

Analogously, in the presence of auxiliary information on $Q_y$, as described in section 6.2, a definition of $\Xi_\varepsilon$ and $H_\varepsilon$, similar to that for $\Omega_\varepsilon$ and $G_\varepsilon$ above, allows a similar conclusion to be reached, since the asymptotic variance matrix $(H_\varepsilon'\Xi_\varepsilon^{-1}H_\varepsilon)^{-1}$ of the ML estimator converges to $(H'\Xi^{-1}H)^{-1}$.

## Appendix C: Response Covariate Dependence

The conditional independence assumptions of section 2.3 may be weakened to allow a dependence on finite partitions of the covariate sample space $\mathcal{X}$. Let $\mathcal{X}_j^{\mathcal{J}}$, $j\in\mathcal{J}$, where $\mathcal{J}$ is finite, be a partition of $\mathcal{X}$, i.e., $\mathcal{X}_j^{\mathcal{J}}\cap\mathcal{X}_l^{\mathcal{J}} = \emptyset$, $j\neq l$, and $\mathcal{X} = \cup_{j\in\mathcal{J}}\mathcal{X}_j^{\mathcal{J}}$. Define the random variable $J = j$ if $X\in\mathcal{X}_j^{\mathcal{J}}$, $j\in\mathcal{J}$. To incorporate dependence on covariates $X$, Assumption 2.3 is modified to

**Assumption C.1** *(Conditional Probability of Observing $Y$.) Observation of $Y$ is conditionally independent of $X$ given $Y$; i.e.,*

$$\begin{aligned} P_{y,j} &= \mathcal{P}\{I^{\mathcal{Y}} = 1|Y = y, X = x\} \\ &= \mathcal{P}\{I^{\mathcal{Y}} = 1|Y = y, J = j\}, \end{aligned}$$

*where $0 < P_{y,j} < 1$, $j\in\mathcal{J}$, $y\in\mathcal{Y}$, $x\in\mathcal{X}$.*

Assumption 2.4 may be altered in a similar fashion, i.e., $G_{y,j} = \mathcal{P}\{I^{\mathcal{X}} = 1|I^{\mathcal{Y}} = 1, Y = y, J = j\}$ and $G_j^{\mathcal{X}} = \mathcal{P}\{I^{\mathcal{X}} = 1|I^{\mathcal{Y}} = 0, Y = y, J = j\}$ where $0 < G_{y,j} < 1$, $0 < G_j^{\mathcal{X}} < 1$, $j\in\mathcal{J}$, $y\in\mathcal{Y}$, $x\in\mathcal{X}$. The choice of identical partition of $\mathcal{X}$ for $P_{y,j}$, $G_{y,j}$ and $G_j^{\mathcal{X}}$ may be relaxed straightforwardly but at the expense of a more involved notation.

Therefore, from Assumptions 2.2, 2.3, and C.1, $\mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1|Y = y, X = x, S = 0\} = P_{y,j}G_y$, $\mathcal{P}\{I^{\mathcal{Y}} = 0, I^{\mathcal{X}} = 1|Y = y, X = x, S = 0\} = (1 - P_{y,j})G^{\mathcal{X}}$, $X\in\mathcal{X}_j^{\mathcal{J}}$, $j\in\mathcal{J}$, etc., i.e., the conditional probabilities of observing respondent, INR($y$), INR($x$) and INR units given $(Y, X)$ now display a discrete dependence on $X$ in addition to that on $Y$.

Mirroring (2.6) in section 2.3, $\mathcal{P}\{I^{\mathcal{Y}} = 1, I^{\mathcal{X}} = 1\} = \sum_{y\in\mathcal{Y}}\sum_{j\in\mathcal{J}}P_{y,j}G_yQ_{y,j}$, where $Q_{y,j} = \mathcal{P}\{Y = y, J = j\} = \int_{\mathcal{X}_j^J}\mathcal{P}\{y|x,\theta\}f_X(x)dx$, cf. (2.2). Similarly to (2.7) and (2.8), define $H_{y,j} = \mathcal{P}\{Y = y, J = j, I^{\mathcal{Y}} = 1, S = 0\}$ with a similar definition for $H_{y,j}^{nr}$. Then $Q_{y,j} = (H_{y,j} + H_{y,j}^{nr})/(1 - H_s)$ and $P_{y,j} = H_{y,j}/Q_{y.j}(1 - H_s)$. Cf. (2.9) and (2.10).

Define the binary indicators $I_j^{\mathcal{J}} = 1$, if $X\in\mathcal{X}_j^{\mathcal{J}}$, and 0 otherwise, $j\in\mathcal{J}$. Therefore, under Assumptions 2.1-2.3 and C.1, the joint sample density function of $Y$, $X$, $I^{\mathcal{Y}}$, $I^{\mathcal{X}}$, $\{I_j^{\mathcal{J}}\}_{j\in\mathcal{J}}$ and $S$

becomes

$$h(y,x,i^{\mathcal{Y}},i^{\mathcal{X}},\{i_j^{\mathcal{J}}\}_{j\in\mathcal{J}},s) = \left\{ \left[ \prod_{j\in\mathcal{J}} \left( \frac{H_{y,j}}{Q_{y,j}} G_y \mathcal{P}\{y|x,\theta\} f_X(x) \right)^{i^{\mathcal{X}} i_j^{\mathcal{J}}} \left( \sum_{j\in\mathcal{J}} H_{y,j}(1-G_y) \right)^{(1-i^{\mathcal{X}})} \right]^{i^{\mathcal{Y}}} \right.$$
$$\times \left[ \prod_{j\in\mathcal{J}} \left( (1-H_s - \sum_{v\in\mathcal{Y}} \frac{H_{v,j}}{Q_{v,j}} \mathcal{P}\{v|x,\theta\}) G^{\mathcal{X}} f_X(x) \right)^{i^{\mathcal{X}} i_j^{\mathcal{J}}} \right.$$
$$\left. \left. \times \left( \sum_{j\in\mathcal{J}} \left( (1-H_s) \sum_{v\in\mathcal{Y}} Q_{v,j} - \sum_{v\in\mathcal{Y}} H_{v,j} \right) (1-G^{\mathcal{X}}) \right)^{(1-i^{\mathcal{X}})} \right]^{(1-i^{\mathcal{Y}})} \right\}^{1-s}$$
$$\times [H_s f_X(x)]^s.$$

Essentially, $H_y$ and $Q_y$ are replaced by $H_{y,j}$ and $Q_{y,j}$ respectively and integration over $\mathcal{X}$ ($\int_{\mathcal{X}}$) by summation over $j \in \mathcal{J}$ and integration over $\mathcal{X}_j$ ($\sum_{j\in\mathcal{J}} \int_{\mathcal{X}_j}$). If Assumption 2.4 is also relaxed as outlined above, $G_{y,j}$ and $G_j^{\mathcal{X}}$ are substituted for $G_y$ and $G^{\mathcal{X}}$ respectively

Analysis proceeds as in section 4 with the inclusion of these additional parameters.

## References


Allison, P.D. (2001): *Missing Data.* Sage University Papers Series on Quantitative Applications in the Social Sciences.

Basu, D. (1977): "On the Elimination of Nuisance Parameters", *Journal of the American Statistical Association*, 72, 355-366.

Cameron, A.C. and Trivedi, P.K. (2005): *Microeconometrics.* Cambridge: Cambridge University Press.

Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995): *Measurement Error in Nonlinear Models.* New York: Chapman and Hall.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics,* 34, 305-334.

Cosslett, S.R. (1981a): "Efficient Estimation of Discrete-Choice models", in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, 51-111. Cambridge: MIT Press.

Cosslett, S.R. (1981b): "Maximum Likelihood Estimators for Choice-Based Samples", *Econometrica*, 49, 1289-1316.

Cosslett, S.R. (1993): "Estimation from Endogenously Stratified Samples", in G.S. Maddala, C.R. Rao and H.D. Vinod (eds.), *Handbook of Statistics,* Volume *11*, 1-43. Amsterdam: Elsevier.

Cosslett, S.R. (1997): "Nonparametric Maximum Likelihood Methods", in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics,* Volume *15*, 385-404. Amsterdam: Elsevier.

Durrant, G.B. and Skinner, C. (2006): "Using Data Augmentation to Correct for Non-Ignorable Non-Response When Surrogate Data are Available: An Application to the Distribution of Hourly Pay", *Journal of the Royal Statistical Society*, Series A, 169, 605-623.

Engle, R. F., Hendry, D. F. and Richard, J.-F. (1983): "Exogeneity", *Econometrica*, 51, 277-304.

Fitzgerald, J., Gottschalk, P. and Moffit, R. (1997): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics", *Journal of Human Resourses*, 33, 251-299.

Hansen, L.P., Heaton, J. and Yaron, A. (1996): "Finite-Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics*, 14, 262-280.

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", *Annals of Economic and Social Measurement*, 5, 475-492.

Hellerstein, J.K. and Imbens, G.W. (1999): "Imposing Moment Restrictions from Auxiliary Data by Weighting", *Review of Economics and Statistics 81, 1-14.*

Hirano, K., Imbens, G.W., Ridder, G. and Rubin, D.B. (1998): "Combining Panel Data Sets with Attrition and Refreshment Samples", *Econometrica*, 69, 1645-1659.

Horowitz, J.L. and Manski, C.F. (1995): "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica*, 63, 281-302.

Horowitz, J.L. and Manski, C.F. (1998): "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations", *Journal of Econometrics*, 84, 37-58.

Horowitz, J.L. and Manski, C.F. (2001): "Imprecise Identification from Incomplete Data". Working paper, Northwestern University.

Hsieh, D.A., Manski, C.F. and McFadden, D. (1985): "Estimation of Response Probabilities from Augmented Retrospective Observations", *Journal of the American Statistical Association,* 80, 651-662.

Hu, Y. and Schennach, S. (2008): "Identification and Estimation of Nonclassical Nonlinear Errors-In-Variables Models with Continuous Distributions using Instruments", *Econometrica*, 76, 195-216.

Imbens, G.W. (1992): "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling", *Econometrica*, 60, 1187-1214.

Imbens, G.W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies,* 64, 359-383.

Imbens, G.W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review", *Review of Economics and Statistics*, 86, 4-29.

Imbens, G.W. and Lancaster, T. (1994): "Combining Micro and Macro Data in Microeconometric Models", *Review of Economic Studies*, 61, 655-680.

Imbens, G.W. and Lancaster, T. (1996): "Efficient Estimation and Stratified Sampling", *Journal of Econometrics*, 74, 289-318.

Imbens, G.W., R.H. Spady and Johnson, P. (1998): "Information Theoretic Approaches to Inference in Moment Condition Models", *Econometrica,* 66, 333-357.

Kitamura, Y., and Stutzer, M. (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.

Lancaster, T. and Imbens, G.W. (1996): "Case-Control Studies with Contaminated Controls", *Journal of Econometrics*, 71, 145-160.

Lawless, J.F., Kalbfleisch, J.D. and Wild, C.J (1999): "Semiparametric Methods for Response-Selective and Missing Data Problems in Regression", *Journal of the Royal Statistical Society*, Series B, 61, 413-438.

Li, G. and Qin, J. (1998): "Semiparametric Likelihood-Based Inference for Biased and Truncated Data When the Total Sample Size is Known", *Journal of the Royal Statistical Society*, Series B, 60, 243-254.

Little, R.J.A. and Rubin, D.B. (2002): *Statistical Analysis with Missing Data*. (Second edition). Hoboken, New Jersey: Wiley.

Manski, C. and Lerman, S. (1977): "The Estimation of Choice Probabilities from Choice Based Samples", *Econometrica*, 45, 1977-1988.

Newey, W. and McFadden, D. (1994): "Large Sample Estimation and Hypothesis Testing", in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, 2111-2245. Amsterdam: Elsevier.

Newey, W.K., and Smith, R.J. (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood." *Econometrica*, 72, 219-255.

Newey, W.K. and West, K.D. (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777-787.

Owen, A. (2001): *Empirical Likelihood*. New York: Chapman and Hall.

Qin, J. and Lawless, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300-325.

Ramalho, E.A., and Smith, R.J. (2003): "Discrete Choice Nonresponse." CWP 07/03, Centre for Microdata Methods and Practice, I.F.S. and U.C.L.. http://cemmap.ifs.org.uk/wps/cwp0307.pdf

Ridder, G. (1990): "Attrition in Multi-Wave Panel Data", in J. Hartog, G. Ridder and J. Theeuwes (eds.), *Panel Data and Labour Market Studies*, Elsevier Science Publishers, North-Holland, 45-68.

Rubin, D.B. (1976): "Inference and Missing Data", *Biometrika*, 63(3), 581-592.

Schafer, J.L. (1997): *Analysis of Incomplete Multivariate Data*. New York: Chapman and Hall.

Skinner, C., Stuttard, N., Beissel-Durrant, G. and Jenkins, J. (2002): "The Measurement of Low Pay in the UK Labour Force Survey", *Oxford Bulletin of Economics and Statistics*, 64, 653-676.

Smith, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation", *Economic Journal*, 107, 503-519.

Smith, R. J. (2001): "GEL Methods for Moment Condition Models." Revised version CWP 19/04, cemmap, I.F.S. and U.C.L., available at http://cemmap.ifs.org.uk/wps/cwp0419.pdf. Forthcoming *Econometric Theory*.

Stewart, M.B. (1983): "On Least Squares Estimation when the Dependent Variable is Grouped", *Review of Economic Studies*, 50, 737-753.

Tang, G., Little, R.J.A., and Raghunathan, T.E. (2003): "Analysis of Multivariate Missing Data with Nonignorable Nonresponse", *Biometrika*, 90, 747-764.

Weinberg, C.R. and Wacholder, S. (1993): "Prospective Analysis to Case-Control Data under General Multiplicative-Intercept Risk Models", *Biometrika*, 80, 461-465.

Wooldridge, J.M. (1999): "Asymptotic Properties of Weighted M-Estimators for Variable Probability Samples", *Econometrica*, 67, 1385-1406.

Wooldridge, J.M. (2001): "Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples", *Econometric Theory*, 17, 451-470.

Wooldridge, J.M. (2009): *Introductory Econometrics: A Modern Approach*. (Fourth edition.) Boston: South-Western.

Table 1: Ordered Probit Model: Classes

| Classes | $Y = 1$ | $Y = 2$ | $Y = 3$ | $Y = 4$ |
|---|---|---|---|---|
| 2 | $0 \leq \ln(hrrate) \leq \ln(3.6)$ | $\ln(3.6) < \ln(hrrate)$ | | |
| 3 | $0 \leq \ln(hrrate) \leq \ln(3.6)$ | $\ln(3.6) < \ln(hrrate) \leq \ln(5.0)$ | $\ln(5.0) < \ln(hrrate)$ | |
| 4 | $0 \leq \ln(hrrate) \leq \ln(3.6)$ | $\ln(3.6) < \ln(hrrate) \leq \ln(4.0)$ | $\ln(4.0) < \ln(hrrate) \leq \ln(5.0)$ | $\ln(5.0) < \ln(hrrate)$ |

Table 2: Ordered Probit: RS ML and INR GMM Estimation Results

| | RS | | INR | |
|---|---|---|---|---|
| | $C = 3$ | $C = 4$ | $C = 3$ | $C = 4$ |
| Intercept | 1.493*** | 1.480*** | 2.311*** | 2.293*** |
| | (0.042) | (0.042) | (0.118) | (0.119) |
| Months employed | 0.016*** | 0.016*** | 0.027*** | 0.027*** |
| | (0.002) | (0.002) | (0.003) | (0.003) |
| Part-time | -0.307*** | -0.309*** | -0.667*** | -0.671*** |
| | (0.026) | (0.026) | (0.063) | (0.063) |
| Occupation | | | | |
| Managers, admin, professional, associate prof. | 0.464*** | 0.472*** | 1.152*** | 1.159*** |
| | (0.035) | (0.035) | (0.067) | (0.067) |
| Craft and related | 0.545*** | 0.546*** | 0.447*** | 0.452*** |
| | (0.051) | (0.051) | (0.086) | (0.086) |
| Clerical and secretarial | -0.036 | -0.035 | 0.176* | 0.176* |
| | (0.031) | (0.030) | (0.093) | (0.093) |
| Head of houshold | 0.201*** | 0.208*** | 0.194*** | 0.201*** |
| | (0.028) | (0.027) | (0.056) | (0.057) |
| Married | 0.163*** | 0.167*** | 0.216*** | 0.220*** |
| | (0.026) | (0.026) | (0.052) | (0.052) |
| Qualifications | | | | |
| Degree level | 0.214*** | 0.215*** | 0.464*** | 0.464*** |
| | (0.072) | (0.071) | (0.075) | (0.076) |
| NVQ level 1/equiv | -0.145*** | -0.143*** | -0.373*** | -0.372*** |
| | (0.030) | (0.030) | (0.065) | (0.065) |
| None | -0.320*** | -0.328*** | -0.607*** | -0.617*** |
| | (0.029) | (0.029) | (0.084) | (0.084) |
| Pay period less than weekly | -0.323* | -0.345** | -1.214*** | -1.230*** |
| | (0.166) | (0.164) | (0.341) | (0.326) |
| Size (25+ employees at workplace) | 0.220*** | 0.233*** | 0.162*** | 0.173*** |
| | (0.024) | (0.024) | (0.056) | (0.056) |
| Industry | | | | |
| Distribution, hotels & restaurants | -0.223*** | -0.230*** | -0.375*** | -0.384*** |
| | (0.026) | (0.026) | (0.071) | (0.071) |
| Industry: other services | -0.301*** | -0.306*** | -0.477*** | -0.481*** |
| | (0.049) | (0.049) | (0.110) | (0.110) |
| Region: London | 0.327*** | 0.336*** | 0.543*** | 0.551*** |
| | (0.057) | (0.057) | (0.095) | (0.096) |

Notes: Estimated standard errors in parentheses; ***, ** and * denote significance at the 0.01, 0.05 and 0.10 levels respectively.

Table 3: Experimental Designs: Missing Data Patterns

| Models | $C = 2$ | | | | $C = 3$ | | | | | $C = 4$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Designs | $a$ | $b$ | $c$ | $d$ | $a$ | $b$ | $c$ | $d$ | $e$ | $a$ | $b$ | $c$ | $d$ | $e$ |
| $P_1$ | 1.0 | 0.50 | 0.98 | 0.98 | 1.0 | 0.50 | 0.98 | 0.98 | 0.98 | 1.0 | 0.50 | 0.98 | 0.98 | 0.98 |
| $P_2$ | 1.0 | 0.50 | 0.50 | 0.70 | 1.0 | 0.50 | 0.50 | 0.70 | 0.70 | 1.0 | 0.50 | 0.98 | 0.98 | 0.80 |
| $P_3$ | | | | | 1.0 | 0.50 | 0.50 | 0.70 | 0.50 | 1.0 | 0.50 | 0.50 | 0.70 | 0.70 |
| $P_4$ | | | | | | | | | | 1.0 | 0.50 | 0.50 | 0.70 | 0.50 |

Table 4: Ordered Probit: $C = 2$

| $N$ | Estimator | $\theta_0$ | | | | $\theta_1$ | | | | $\theta_2$ | | | | $\theta_3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE |
| | | Mean | Median | | | Mean | Median | | | Mean | Median | | | Mean | Median | | |
| | | Design $a$: $P_y = (1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 500 | RS | .004 | .003 | .165 | .132 | .093 | .048 | .014 | .011 | .017 | .019 | .193 | .154 | .028 | .004 | .270 | .196 |
| | | Design $b$: $P_y = (0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .008 | .004 | .254 | .178 | .202 | .070 | .022 | .017 | .055 | .039 | .314 | .224 | .033 | -.004 | .473 | .280 |
| | | Design $c$: $P_y = (0.98, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.161 | -.164 | .194 | .374 | .183 | .141 | .016 | .013 | .105 | .099 | .230 | .190 | .126 | .096 | .348 | .249 |
| | INR | .008 | .006 | .159 | .123 | .088 | .043 | .014 | .011 | .018 | .023 | .198 | .158 | .048 | .006 | .611 | .216 |
| | UNR | .074 | .049 | .231 | .202 | .048 | .014 | .014 | .011 | -.006 | -.017 | .208 | .164 | .067 | -.022 | .687 | .282 |
| | | Design $d$: $P_y = (0.98, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.076 | -.078 | .181 | .208 | .138 | .086 | .015 | .012 | .061 | .066 | .210 | .170 | .068 | .046 | .260 | .210 |
| | INR | .007 | .005 | .160 | .124 | .083 | .045 | .014 | .011 | .017 | .025 | .197 | .157 | .034 | .005 | .535 | .201 |
| | UNR | .074 | .037 | .247 | .204 | .043 | .012 | .013 | .010 | -.014 | -.010 | .196 | .156 | .101 | -.031 | 1.277 | .321 |
| | | Design $a$: $P_y = (1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 1000 | RS | .001 | .000 | .121 | .095 | .055 | .045 | .009 | .007 | .011 | .007 | .139 | .111 | .014 | .005 | .177 | .139 |
| | | Design $b$: $P_y = (0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .001 | .000 | .154 | .122 | .091 | .060 | .013 | .011 | .016 | .017 | .195 | .155 | .029 | .013 | .247 | .194 |
| | | Design $c$: $P_y = (0.98, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.166 | -.167 | .138 | .380 | .146 | .137 | .011 | .009 | .084 | .077 | .163 | .137 | .104 | .098 | .197 | .180 |
| | INR | .004 | .006 | .112 | .087 | .051 | .038 | .010 | .007 | .010 | .009 | .145 | .116 | .012 | .000 | .169 | .132 |
| | UNR | .047 | .029 | .175 | .138 | .034 | .026 | .010 | .007 | -.012 | -.013 | .149 | .120 | -.009 | -.013 | .176 | .140 |
| | | Design $d$: $P_y = (0.98, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.081 | -.083 | .129 | .196 | .101 | .089 | .010 | .008 | .049 | .043 | .151 | .123 | .060 | .051 | .187 | .155 |
| | INR | .005 | .003 | .111 | .086 | .044 | .034 | .010 | .007 | .013 | .007 | .149 | .116 | .012 | .001 | .174 | .134 |
| | UNR | .050 | .016 | .194 | .144 | .031 | .020 | .009 | .007 | -.015 | -.020 | .144 | .115 | -.013 | -.021 | .177 | .142 |

Table 5: Ordered Probit: $C = 3$

| $N$ | Estimator | $\theta_0$ | | | | $\theta_1$ | | | | $\theta_2$ | | | | $\theta_3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE |
| | | Mean | Median | | | Mean | Median | | | Mean | Median | | | Mean | Median | | |
| | | Design $a$: $P_y = (1.00, 1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 500 | RS | .003 | .004 | .144 | .114 | .067 | .033 | .011 | .009 | .011 | .012 | .163 | .131 | .017 | .008 | .200 | .154 |
| | | Design $b$: $P_y = (0.50, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .003 | .003 | .191 | .150 | .130 | .054 | .016 | .013 | .031 | .028 | .239 | .189 | .035 | .021 | .294 | .222 |
| | | Design $c$: $P_y = (0.98, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.135 | -.134 | .181 | .316 | .193 | .165 | .014 | .012 | .130 | .124 | .212 | .181 | .141 | .125 | .249 | .232 |
| | INR | .024 | .022 | .149 | .124 | .047 | .024 | .012 | .009 | -.013 | -.013 | .183 | .145 | .018 | -.008 | .422 | .178 |
| | | Design $d$: $P_y = (0.98, 0.70, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.062 | -.063 | .163 | .177 | .130 | .098 | .012 | .010 | .074 | .072 | .188 | .155 | .086 | .066 | .254 | .192 |
| | INR | .017 | .015 | .143 | .114 | .049 | .026 | .011 | .009 | -.004 | -.003 | .172 | .137 | .013 | -.007 | .231 | .159 |
| | | Design $e$: $P_y = (0.98, 0.70, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.145 | -.146 | .173 | .336 | .150 | .120 | .013 | .011 | .091 | .087 | .202 | .167 | .114 | .104 | .241 | .211 |
| | INR | .019 | .019 | .146 | .119 | .048 | .029 | .011 | .009 | -.003 | -.006 | .177 | .141 | .008 | -.005 | .205 | .156 |
| | | Design $a$: $P_y = (1.00, 1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 1000 | RS | .000 | -.001 | .102 | .082 | .034 | .028 | .007 | .006 | .008 | .008 | .119 | .093 | .007 | .001 | .136 | .107 |
| | | Design $b$: $P_y = (0.50, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .000 | -.001 | .132 | .105 | .048 | .030 | .010 | .008 | .007 | .000 | .167 | .134 | .016 | .007 | .199 | .154 |
| | | Design $d$: $P_y = (0.98, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.138 | -.139 | .128 | .318 | .152 | .146 | .009 | .008 | .112 | .104 | .152 | .134 | .125 | .115 | .174 | .180 |
| | INR | .018 | .017 | .107 | .091 | .013 | .002 | .008 | .006 | -.014 | -.018 | .133 | .106 | -.009 | -.016 | .140 | .110 |
| | | Design $d$: $P_y = (0.98, 0.70, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.066 | -.066 | .115 | .163 | .097 | .086 | .008 | .007 | .061 | .060 | .134 | .111 | .066 | .058 | .155 | .136 |
| | INR | .011 | .011 | .098 | .081 | .021 | .013 | .007 | .006 | -.009 | -.008 | .128 | .101 | -.005 | -.011 | .136 | .107 |
| | | Design $e$: $P_y = (0.98, 0.70, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.149 | -.149 | .122 | .342 | .121 | .116 | .009 | .007 | .071 | .070 | .144 | .120 | .095 | .088 | .163 | .155 |
| | INR | .014 | .013 | .105 | .086 | .022 | .011 | .008 | .006 | -.013 | -.016 | .130 | .102 | -.008 | -.011 | .133 | .105 |

Table 6: Ordered Probit: $C = 4$

| $N$ | Estimator | $\theta_0$ | | | | $\theta_1$ | | | | $\theta_2$ | | | | $\theta_3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE | Bias | | SD | MAE |
| | | Mean | Median | | | Mean | Median | | | Mean | Median | | | Mean | Median | | |
| | | Design $a$: $P_y = (1.00, 1.00, 1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 500 | RS | .003 | .003 | .144 | .114 | .067 | .034 | .011 | .009 | .012 | .013 | .163 | .131 | .017 | .008 | .200 | .154 |
| | | Design $b$: $P_y = (0.50, 0.50, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .003 | .004 | .193 | .151 | .126 | .059 | .016 | .013 | .032 | .030 | .241 | .192 | .034 | .014 | .295 | .222 |
| | | Design $c$: $P_y = (0.98, 0.98, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.143 | -.144 | .174 | .333 | .164 | .128 | .013 | .011 | .104 | .099 | .203 | .170 | .125 | .107 | .244 | .217 |
| | INR | .020 | .020 | .144 | .118 | .045 | .018 | .011 | .009 | -.006 | -.002 | .176 | .141 | .014 | -.014 | .324 | .168 |
| | | Design $d$: $P_y = (0.98, 0.98, 0.70, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.067 | -.067 | .160 | .183 | .121 | .086 | .012 | .010 | .063 | .060 | .184 | .151 | .075 | .063 | .225 | .182 |
| | INR | .014 | .013 | .141 | .111 | .059 | .025 | .011 | .009 | .003 | .007 | .173 | .137 | .012 | -.006 | .207 | .156 |
| | | Design $e$: $P_y = (0.98, 0.80, 0.70, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.146 | -.147 | .171 | .339 | .147 | .122 | .013 | .011 | .087 | .089 | .200 | .166 | .111 | .102 | .240 | .210 |
| | INR | .018 | .018 | .145 | .116 | .052 | .031 | .011 | .009 | .000 | .000 | .175 | .140 | .017 | -.006 | .320 | .166 |
| | | Design $a$: $P_y = (1.00, 1.00, 1.00, 1.00)$ | | | | | | | | | | | | | | | |
| 1000 | RS | .000 | .000 | .102 | .082 | .034 | .028 | .007 | .006 | .008 | .008 | .119 | .093 | .007 | .001 | .136 | .108 |
| | | Design $b$: $P_y = (0.50, 0.50, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | .000 | .000 | .131 | .105 | .049 | .039 | .010 | .008 | .006 | -.004 | .166 | .133 | .016 | .006 | .199 | .155 |
| | | Design $c$: $P_y = (0.98, 0.98, 0.50, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.148 | -.148 | .125 | .339 | .133 | .126 | .009 | .008 | .085 | .085 | .147 | .125 | .109 | .100 | .168 | .167 |
| | INR | .014 | .013 | .102 | .084 | .020 | .007 | .007 | .006 | -.009 | -.016 | .132 | .104 | -.004 | -.007 | .137 | .108 |
| | | Design $d$: $P_y = (0.98, 0.98, 0.70, 0.70)$ | | | | | | | | | | | | | | | |
| | RS | -.072 | -.072 | .113 | .173 | .089 | .082 | .008 | .007 | .050 | .048 | .132 | .108 | .059 | .049 | .152 | .131 |
| | INR | .010 | .010 | .098 | .080 | .025 | .020 | .007 | .006 | -.004 | -.007 | .129 | .101 | -.004 | -.012 | .135 | .106 |
| | | Design $e$: $P_y = (0.98, 0.80, 0.70, 0.50)$ | | | | | | | | | | | | | | | |
| | RS | -.150 | -.150 | .121 | .345 | .115 | .114 | .009 | .007 | .066 | .066 | .142 | .118 | .091 | .086 | .162 | .153 |
| | INR | .013 | .013 | .103 | .085 | .017 | .007 | .007 | .006 | -.010 | -.015 | .129 | .101 | -.007 | -.011 | .133 | .105 |

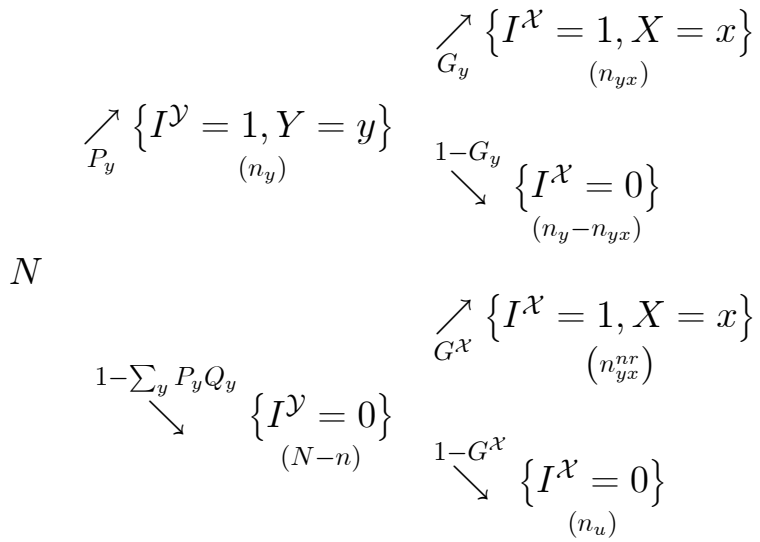
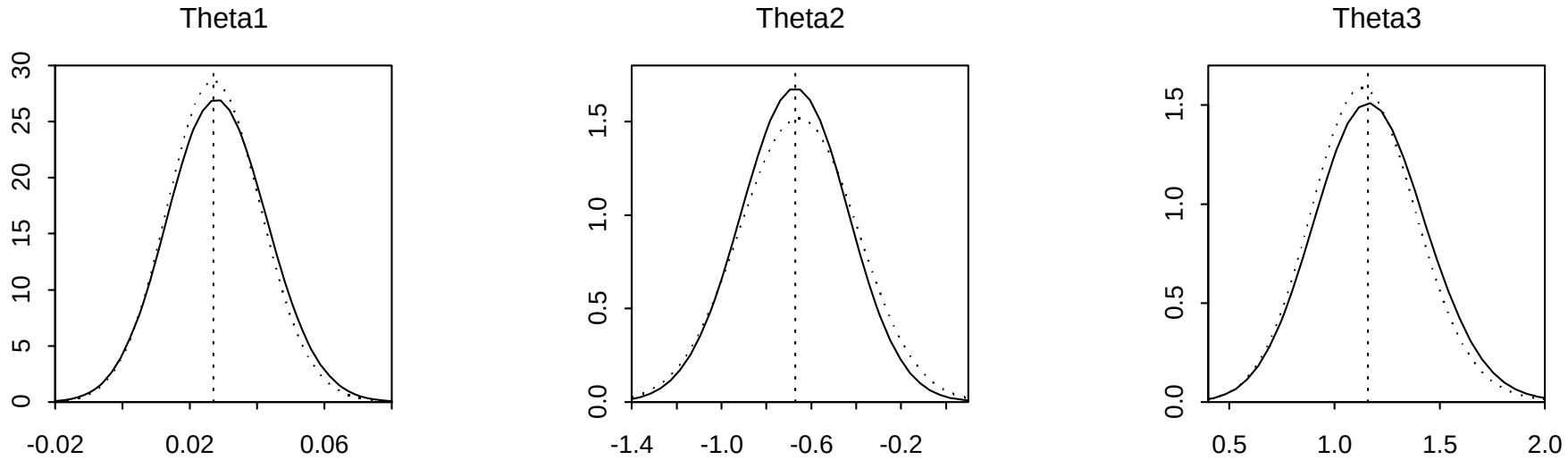


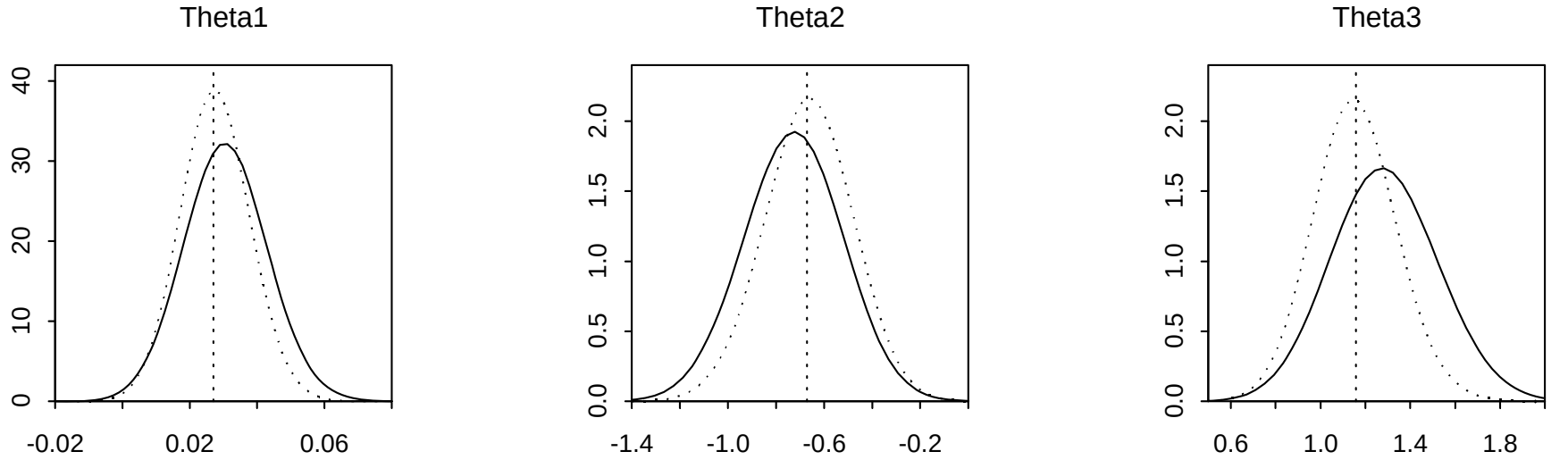

**Figure 1: Missingness Structure**

FIGURE 2: ESTIMATED SAMPLING DENSITIES: C=4, N=1000

I) Design b: P = (0.50,0.50,0.50,0.50)



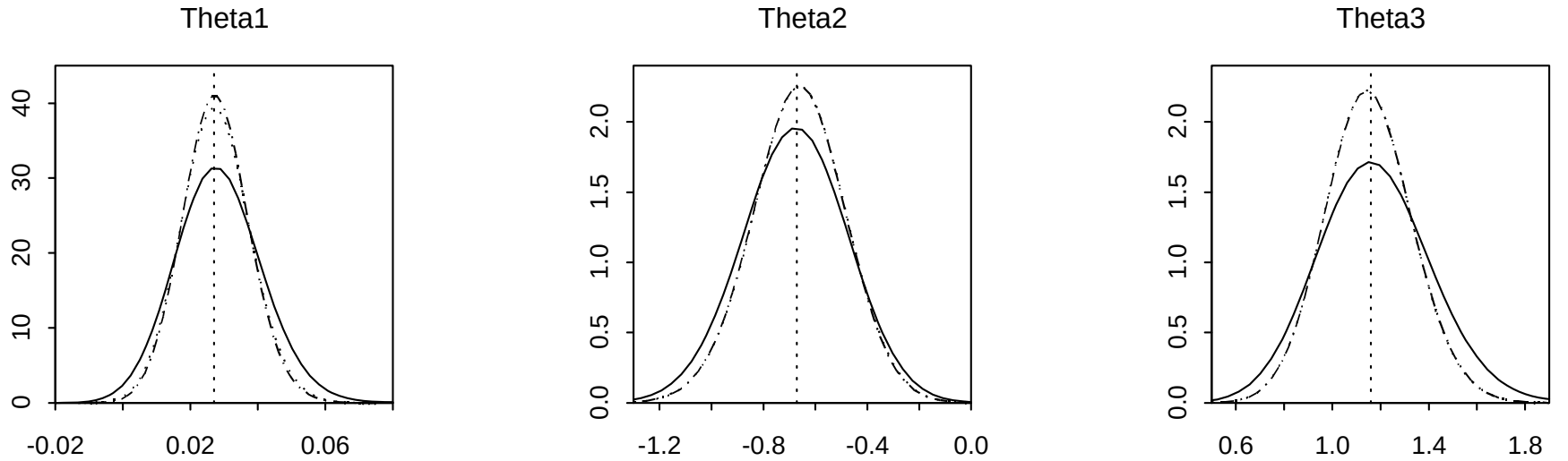

II) Design c: P = (0.98,0.98,0.50,0.50)



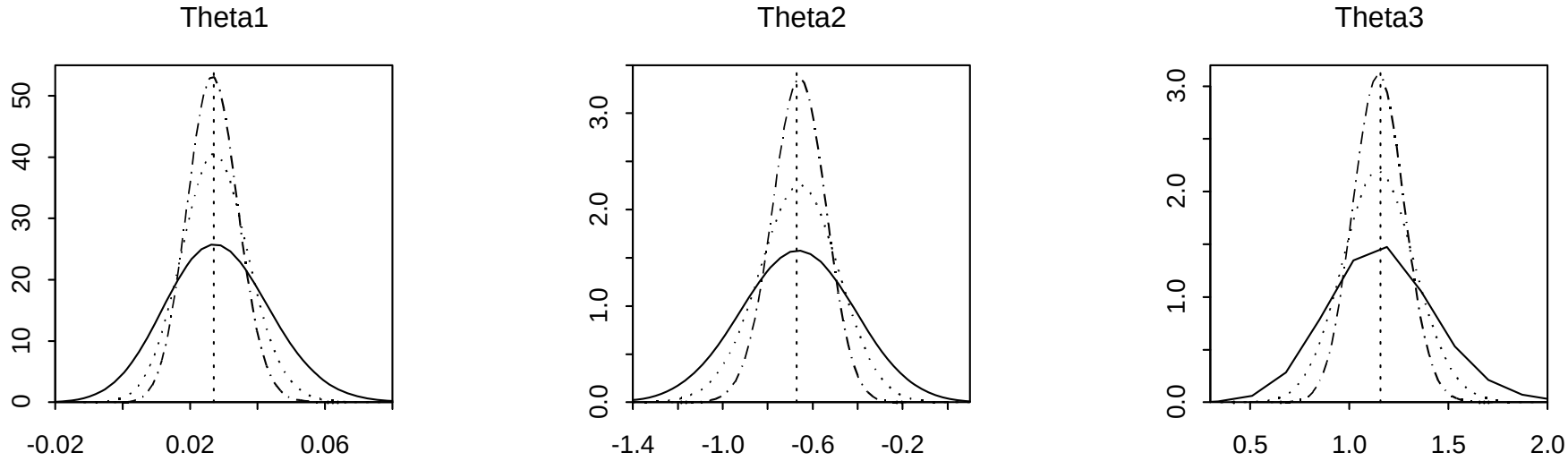

Notes: RS (solid line) and INR (dashed line) estimators. The vertical dotted line indicates the true value of the parameter.

FIGURE 3: ESTIMATED SAMPLING DENSITIES: INR ESTIMATORS

I) Design e: P = (0.98,0.80,0.70,0.50)




II) Design e: P = (0.98,0.80,0.70,0.50)




Note I): N=1000: C=2 (solid line), C=3 (dotted line), and C=4 (dotted-dashed line). The vertical dotted line indicates the true value of the parameter.

Note II): C=4: N=500 (solid line), N=1000 (dotted line), and N=2000 (dotted-dashed line). The vertical dotted line indicates the true value of the parameter.