Uses of Probabilistic Models of Unit Nonresponse

G. Rohwer

Version 1 **April 2011**

Contents

1 Introduction

Attempts to cope with unit nonresponse in surveys are often based on probabilistic models which posit 'probabilities of response', and then think of these probabilities as being dependent on a set of identifiable (and known) variables. This paper discusses how to understand and use such models.

The discussion is based on a distinction between two purposes for which the sampled data can be used. (a) Descriptive estimation of statistical distributions which are defined for a particular target population from which the sample is drawn. (b) Estimation of functional models (models formalizing probabilistic rules) which concern the behavior of a generic unit conditional on known values of some variables. The distinction is helpful because relationships with response models are different. Being itself a kind of functional model, a response model can be directly integrated into the primarily interesting functional model in order to assess what can be estimated with the available data. In contrast, there is no easy way to integrate response models into the standard approach to descriptive estimation that is based on randomization via a sampling design.

The paper is only concerned with unit nonresponse resulting from decisions of (through the sampling design) selected units after they have been contacted. It will be assumed throughout that respondents provide complete information about all variables of interest. In section 2 I discuss descriptive estimation; functional models will be considered in section 3. Section 4 concludes with a suggestion for the presentation of data.

2 Consideration of Descriptive Estimation

2.1 The Formal Framework

Let Ω denote the target population, a finite set of units. The interest may concern:

a) The distribution of a statistical variable Y, to be understood as a function $Y : \Omega \longrightarrow \mathcal{Y}$ which assigns to each unit $\omega \in \Omega$ an element $Y(\omega)$ of the variable's property space $\mathcal{Y}$. (Y, like all other variables introduced below, may consist of two or more components.) The distribution of Y in Ω will be denoted by $\mathrm{P}[Y]$; for specific values I use the notation $\mathrm{P}(Y=y)$, meaning the proportion of units in Ω having the value y of the variable Y. Of course, one might also be interested in quantities derived from Y's distribution (e.g., the mean of Y).

b) Regression functions which are derived from the distribution of a two-dimensional statistical variable $(X,Y):\Omega \longrightarrow \mathcal{X}\times\mathcal{Y}$. I use the notation

$$x \longrightarrow \mathrm{P}[Y \mid X=x] \tag{1}$$

meaning that the regression function assigns to each value $x\in\mathcal{X}$ the conditional distribution of Y given $X=x$. Specific values of the conditional distribution will be denoted by $\mathrm{P}(Y=y \mid X=x)$.[1]

Now let $\mathcal{S}\subset\Omega$ denote a sample of units randomly drawn from Ω. Until further notice I assume that $\mathcal{S}$ is a simple random sample (design-based weights will be considered in section 2.4.4). Variables restricted to the sample will be denoted by Y^s and (X^s,Y^s), respectively. If complete information would be available, one could use $\mathrm{P}[Y^s]$ and $\mathrm{P}[Y^s|X^s=x]$ for estimating $\mathrm{P}[Y]$ and $\mathrm{P}[Y|X=x]$. However, in case of unit nonresponse, one knows values of the variables only for a subset of the sample, say $\mathcal{S}^r\subset\mathcal{S}$. This can be described by introducing a variable $R^s:\mathcal{S}\longrightarrow\{0,1\}$, with $R^s(\omega)=1$ if $\omega\in\mathcal{S}^r$ and $R^s(\omega)=0$ otherwise. The information available is then given by the distributions $\mathrm{P}[Y^s|R^s=1]$ and $\mathrm{P}[Y^s|X^s=x,R^s=1]$, respectively. In case of descriptive estimation, the question is how one can use this information for the estimation of $\mathrm{P}[Y]$ and $\mathrm{P}[Y|X=x]$. More specifically, there are two questions:

(1) Under which conditions does $\mathrm{P}[Y^s|R^s=1]$ provide a plausible estimate of $\mathrm{P}[Y^s]$ (that is, nonresponse can be ignored)?

(2) Given that there is a nonresponse bias (defined in some way by quantifying the difference between $\mathrm{P}[Y^s]$ and $\mathrm{P}[Y^s|R^s=1]$), can one find a better estimate of $\mathrm{P}[Y^s]$ which reduces this bias?

[1] For ease of notation, I suppose that all variables have a discrete property space.

Analogous questions can be formulated for conditional distributions (regression functions).

2.2 Probabilistic Models of Unit Nonresponse

The basic idea is to posit 'probabilities of response' and then to think of these probabilities as being dependent, in some regular way, on values of identifiable variables. As a formal framework for this idea, I use the notation

$$h \longrightarrow \Pr(\dot{R}=1 \mid \ddot{H}=h) \tag{2}$$

to be interpreted as a probabilistic rule: *If* the variable $\ddot{H}$ has the value h (an element in the property space $\mathcal{H}$), *then* the probability of response, recorded by $\dot{R}=1$, is $\Pr(\dot{R}=1 \mid \ddot{H}=h)$.[2] Notice that neither $\ddot{H}$ nor $\dot{R}$ are statistical variables.[3] In contrast to the statistical variable R^s, $\dot{R}$ is a random variable (indicated by the dot). Moreover, there is no unconditional distribution of $\dot{R}$; the model only provides probability distributions conditional on values of $\ddot{H}$. The model neither requires nor implies a distribution of $\ddot{H}$; this is an exogenous variable of the model (indicated by two dots), and only serves to specify the *if*-part of the rule formulated by the model (2).

How to understand the random variable $\dot{R}$? A first question concerns the nature of the associated conditional probability distributions. In a first understanding, the model formulates a rule for predictions: Referring to a unit, say ω, being approached to participate in the survey (in order to get information about values of the variables of interest), and knowing ω's value of $\ddot{H}$, one could use (2) to probabilistically predict whether ω will participate (and one will get the information). However, there is no

[2] Pr is used for probabilities and should be distinguished from frequencies, referred to by P, which are defined as proportions in finite reference sets.

[3] I use this term to denote a function having a sample or target population as its domain, like Y introduced above.

point in using (2) for such predictions; and in fact, when dealing with the questions formulated at the end of section 2.1, the most important role is played by assumptions, not entailed in the formulation of the response model. These assumptions concern the approximate independence of $\dot{R}$ from the variables of interest, conditional on values of the variable $\ddot{H}$ that is used in the response model.

A further question concerns the demarcation of units for which the model is intended to hold. I distinguish between a *global* response model, assumed to be valid for all units in the target population, and a *local* response model, assumed to be valid only for units in the selected sample. In the following considerations I always assume a local response model.

2.3 Formulating Independence Assumptions

How to formulate assumptions about conditional independence between $\dot{R}$ and the variables of interest depends on the conceptual framework. In this section, I consider first a direct reference to statistical variables, then very briefly a probabilistic modeling approach. (Another kind of modeling approach which is not restricted to descriptive estimation will be discussed in section 3.)

2.3.1 Direct Use of Statistical Variables

Suppose that the statistical variable of interest is Y. In order to establish a relationship of the response model with this variable, one can include a correspondingly defined exogenous variable, $\ddot{Y}$, in the model; the independence assumption can then be formulated as

$$\Pr(\dot{R}=1 \mid \ddot{H}=h, \ddot{Y}=y) = \Pr(\dot{R}=1 \mid \ddot{H}=h) \tag{3}$$

This assumption can be used to argue that

$$\mathrm{P}(R^s=1 \mid H^s=h, Y^s=y) \;\approx\; \mathrm{P}(R^s=1 \mid H^s=h)$$

(where H^s is a statistical variable, defined for the sample $\mathcal{S}$, corresponding to $\ddot{H}$) holds approximately;[4] and consequently

$$\mathrm{P}(Y^s=y \mid H^s=h) \;\approx\; \mathrm{P}(Y^s=y \mid H^s=h, R^s=1) \tag{4}$$

Finally one can derive

$$\mathrm{P}(Y^s=y) \;\approx\; \sum\nolimits_h \mathrm{P}(Y^s=y \mid H^s=h, R^s=1)\,\mathrm{P}(H^s=h) \tag{5}$$

showing how $\mathrm{P}[Y^s]$, and consequently $\mathrm{P}[Y]$, can plausibly be estimated by using the information from the realized sample $\mathcal{S}^r$ in combination with the distribution of H^s in the complete sample.[5] The argument only requires a local response model. However, in addition to the assumption that all units in the selected sample have a positive response probability, also the actually observed response proportions, $\mathrm{P}(R^s=1|H^s=h)$, must be positive.

An analogous consideration can be used for the estimation of conditional distributions (regression functions). Starting from the independence assumption

$$\Pr(\dot{R}=1 \mid \ddot{H}=h, \ddot{Y}=y, \ddot{X}=x) \;=\; \Pr(\dot{R}=1 \mid \ddot{H}=h, \ddot{X}=x)$$

one could assume that

$$\mathrm{P}(R^s=1 \mid H^s=h, Y^s=y, X^s=x) \;\approx\; \mathrm{P}(R^s=1 \mid H^s=h, X^s=x)$$

is approximately valid for the sample. This would allow one to derive

$$\mathrm{P}(Y^s=y \mid X^s=x, H^s=h, R^s=1) \;\approx\; \mathrm{P}(Y^s=y \mid X^s=x, H^s=h)$$

[4] Using the approximation sign $\approx$ instead of an equal sign is required because the notion of stochastic independence has no direct counterpart for frequencies in finite sets.

[5] If values of H^s are not available for nonrespondents, it might sometimes be possible to use known population proportions $\mathrm{P}(H=h)$ instead of $\mathrm{P}(H^s=h)$. The approach then becomes a form of post-stratification (Holt and Elliot 1991). If H^s consists of several components, say $H^s = (H^s_1, \ldots, H^s_m)$, procedures also depend on whether the complete distribution or only marginal distributions of the components are known. For the latter case, raking procedures have been suggested (Deville, Särndal and Sautory, 1993).

and finally

$$\mathrm{P}(Y^s\!=\!y \mid X^s\!=\!x) \approx \sum\nolimits_h \mathrm{P}(Y^s\!=\!y \mid X^s\!=\!x, H^s\!=\!h, R^s\!=\!1)\, \mathrm{P}(H^s\!=\!h|X^s\!=\!x) \tag{6}$$

One would need the conditional frequencies $\mathrm{P}(H^s = h \mid X^s = x)$ for the complete sample $\mathcal{S}$. If they are not available, an easy solution would be to consider the enlarged regression function $(x, h) \longrightarrow \mathrm{P}[Y \mid X\!=\!x, H\!=\!h]$ which includes H (now with domain Ω) as a regressor variable. More generally formulated, unit nonresponse can be ignored if, conditional on the regressor variables, responses do not depend on the dependent variable of the regression function.

2.3.2 Using a Probabilistic Modeling Approach

The approach just described requires to formally use approximate relationships between statistical distributions which cannot be quantified easily. A probabilistic modeling approach provides an alternative. Supposing that the statistical variable of interest is Y^s, this approach views the values of this variable as realizations of a random variable $\dot{Y}^s$ (having the same property space as Y^s).[6] Using a local response model, the independence assumption can then be formulated with a strict equality sign as

$$\Pr(\dot{R}\!=\!1, \dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h) \;=\; \Pr(\dot{R}\!=\!1 \mid \ddot{H}\!=\!h)\, \Pr(\dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h) \tag{7}$$

This allows one to derive

$$\Pr(\dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h) \;=\; \Pr(\dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h, \dot{R}\!=\!1) \tag{8}$$

corresponding to (4). Based on this equation, one can argue that the observed distributions $\mathrm{P}[Y^s|H^s = h, R^s = 1]$, for $h \in \mathcal{H}$, can be used to estimate the distribution of $\dot{Y}^s$. Since $\dot{Y}^s$ is intended to represent Y^s, one would use

$$\begin{aligned} \Pr(\dot{Y}^s\!=\!y) \;&=\; \sum\nolimits_h \Pr(\dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h, \dot{R}\!=\!1)\, \mathrm{P}(H^s\!=\!h) \\ &\approx \sum\nolimits_h \mathrm{P}(Y^s\!=\!y \mid H^s\!=\!h, R^s\!=\!1)\, \mathrm{P}(H^s\!=\!h) \end{aligned}$$

which employs the observed proportions, $\mathrm{P}(H^s = h)$. Again, the argument presupposes that all units in the selected sample have a positive response probability. (Further requirements depend on the parameterization of $\Pr(\dot{Y}^s\!=\!y \mid \ddot{H}\!=\!h)$ and the chosen estimation method.)

Analogous formulations can be used when the interest concerns conditional distributions (regression functions). This will be further discussed in section 3 where I consider functional models which are not intended to make descriptive statements about a particular target population.

2.3.3 Propensity Scores

The response probabilities posited by a response model can be considered as propensity scores (Rosenbaum and Rubin, 1983). Given a local response model (2), and assuming that there is a variable H^s providing values of $\ddot{H}$ for all units in $\mathcal{S}$, one can define a statistical variable $r_H : \mathcal{S} \longrightarrow \mathbf{R}$ having values

$$r_H(\omega) := \Pr(\dot{R} = 1 \mid \ddot{H} = H^s(\omega)) \tag{9}$$

$r_H(\omega)$ is called ω's propensity score (for response).[7] As discussed by Rosenbaum and Rubin (1983), presupposing the conditional independence (7), propensity scores can be used to construct a coarsening of the conditioning.[8] The argument uses the relationship

$$r_H\!=\!r \iff H^s \in A_r := \{h \in \mathcal{H} \mid \Pr(\dot{R}\!=\!1 \mid \ddot{H}\!=\!h) = r\}$$

[6] In my understanding, this conceptual framework does not require to posit a process that randomly generated the values of Y^s in the selected sample. It suffices to think of the distribution of $\dot{Y}^s$ as a model intended to provide an approximate *representation* of the distribution of Y^s. The suggestion is to conceptually distinguish between 'representation' and 'generation'.

[7] In the present discussion, these propensity scores always concern the conditional probability distribution of the response variable $\dot{R}$. This is different from Rosenbaum and Rubin's discussion where propensity scores concern the assignment to a treatment or a control group. I discuss this difference in section 3.3.

[8] Which degree of coarsening is possible depends on the circumstances. It is quite possible that no coarsening can be achieved; an example will be given in section 2.4.3.

Starting from (7), one can derive[9]

$$\Pr(\dot{R}=1, \dot{Y}^s=y \mid r_H=r) \;=\; \Pr(\dot{R}=1 \mid r_H=r)\, \Pr(\dot{Y}^s=y \mid r_H=r) \quad (10)$$

This shows that it suffices to condition on values of the propensity score variable r_H.

It should be stressed that the notion of propensity scores, in order to become useful, *presupposes* the conditional independence assumption (7) to hold at least approximately. Of course, even without making this assumption, based on knowledge of the values of a variable H^s for all units in the sample $\mathcal{S}$, one can set up a response model in the sense of (2). Depending on the property space $\mathcal{H}$, this might require to employ a parametric form of the model.[10] The model can then be estimated, and propensity scores, as defined in (9), can be calculated for all units in $\mathcal{S}$. However, such an exercise of estimating a response model does not contribute any argument for believing the assumption (7).

2.4 Nonresponse Adjustment Weights

The considerations in section 2.3 concern the first of the two questions distinguished at the end of section 2.1: If the relevant independence assumption holds, unit nonresponse can be ignored. However, such assumptions are not entailed by the response model but must explicitly be added;

[9]The derivation uses the following general rule (α and β are arbitrary suitable expressions): If $\Pr(\alpha \mid \beta, \ddot{H}=h) = c$ for all $h \in A$, then $\Pr(\alpha \mid \beta, \ddot{H} \in A) = c$. Therefore, since for all $h \in A_r$

$$\Pr(\dot{R}=1 \mid \dot{Y}^s=y, \ddot{H}=h) = \Pr(\dot{R}=1 \mid \ddot{H}=h) = r,$$

it follows that $\Pr(\dot{R}=1 \mid \dot{Y}^s=y, \ddot{H} \in A_r) = \Pr(\dot{R}=1 \mid \ddot{H} \in A_r) = r$, or equivalently,

$$\Pr(\dot{R}=1 \mid \dot{Y}^s=y, r_H=r) = \Pr(\dot{R}=1 \mid r_H=r) = r$$

from which (10) immediately follows. Of course, propensity scores must not be zero.

[10]If the number of values of H^s is large, the observed frequencies $\mathrm{P}(R^s=1|H^s=h)$ will often be zero or one and cannot immediately be used as propensity scores intended to be usable as conditions. This can be avoided by defining propensity scores by a parametric model (see Rosenbaum and Rubin, 1983: 47).

and it is difficult (if at all possible) to check and justify such assumptions. It is important, therefore, to consider also the second question: whether, and how, a nonresponse bias that would result from using the realized sample without further adjustment can be reduced. In this section I consider definitions of adjustment weights that possibly contribute to reducing nonresponse bias.

2.4.1 A First Definition of Adjustment Weights

A first definition (corresponding to the argument in section 2.3.1) directly uses reciprocal values of the observed response proportions:

$$w^a_\omega := 1/\mathrm{P}(R^s=1 \mid H^s=H^s(\omega)) \quad (11)$$

and therefore presupposes that these proportions are positive. The definition entails that $\Sigma_{\omega\in\mathcal{S}^r} w^a_\omega = |\mathcal{S}|$. Referring then to 'adjustment cells' $\mathcal{S}_h := \{\omega \in \mathcal{S} \mid H^s(\omega) = h\}$, and using $\mathcal{S}^r_h := \{\omega \in \mathcal{S}_h \mid R^s(\omega) = 1\}$ to denote the subsets of responding units, the approximation (4) can be written as:[11]

$$\sum\nolimits_{\omega\in\mathcal{S}_h} I[Y^s=y](\omega) \;\approx\; \sum\nolimits_{\omega\in\mathcal{S}^r_h} w^a_\omega\, I[Y^s=y](\omega) \quad (12)$$

It follows that

$$\sum\nolimits_{\omega\in\mathcal{S}} I[Y^s=y](\omega) \;\approx\; \sum\nolimits_{\omega\in\mathcal{S}^r} w^a_\omega\, I[Y^s=y](\omega)$$

and this approximation can be used as a starting point for the formulation of estimators which attempt to compensate for unit nonresponse. For example, one can use

$$\frac{1}{|\mathcal{S}|} \sum\nolimits_{\omega\in\mathcal{S}^r} w^a_\omega\, I[Y^s=y](\omega) \quad (13)$$

to estimate $\mathrm{P}(Y^s=y)$.

[11]Here and below I use indicator variables, e.g. $I[Y^s=y](\omega) = 1$ if $Y^s(\omega) = y$, and 0 otherwise.

Table 1 Artificial data for a sample consisting of 10 units.

Unit	Y^s	R^s	H_1^s	H_2^s
ω_1	1	1	1	1
ω_2	1	1	2	2
ω_3	1	1	2	2
ω_4	1	1	2	1
ω_5	1	0	1	2
ω_6	0	1	1	1
ω_7	0	1	1	1
ω_8	0	0	1	2
ω_9	0	0	1	2
ω_{10}	0	0	1	2

2.4.2 Reduction of Nonresponse Bias?

It is well possible that the estimator (13) reduces the nonresponse bias that would result from using the uncorrected estimator $\mathrm{P}(Y^s = y \mid R^s = 1)$, given that the independence assumption (4) does not hold. It is, however, difficult to identify conditions for this to be the case. An example can show that there is no simple relationship with response predictions.

Table 1 shows artificial data for a sample consisting of 10 units. The variable of interest is Y^s, and it is assumed that $\mathrm{P}(Y^s = 1) = 0.5$. The response rate is $\mathrm{P}(R^s = 1) = 0.6$. The uncorrected estimate is obviously biased: $\mathrm{P}(Y^s = 1 | R^s = 1) = 0.67$. The table also shows two auxiliary variables, H_1^s and H_2^s, which can be used to calculate values of the estimator (13). The example shows two things.

a) Using nonresponse adjustment weights can lead to a decrease, but also to an increase of the nonresponse bias. In the example, using H_1^s, the new estimate is 0.53, but using H_2^s, the new estimate is 0.8.

b) H_2^s leads to an increased bias although it provides better predictions of R^s than H_1^s. (Based on H_1^s, the proportion of correct predictions of R^s is 0.7, based on H_2^s, the proportion is 0.8.) This shows that there is no simple relationship between nonresponse bias reduction and the degree to which auxiliary variables allow one to predict responses.

There is, however, a possibly useful argument that starts from the observation that the size of (a version of) nonresponse error depends on the correlation between Y^s and R^s. Using M for the mean of statistical variables, the covariance of Y^s and R^s can be written as

$$\begin{aligned} \mathrm{Cov}(Y^s, R^s) &= \mathrm{M}(Y^s R^s) - \mathrm{M}(Y^s)\,\mathrm{M}(R^s) \\ &= \mathrm{M}(Y^s | R^s = 1)\,\mathrm{M}(R^s) - \mathrm{M}(Y^s)\,\mathrm{M}(R^s) \end{aligned} \tag{14}$$

A version of nonresponse bias is then given by

$$\mathrm{M}(Y^s | R^s = 1) - \mathrm{M}(Y^s) = \frac{\mathrm{Cov}(Y^s, R^s)}{\mathrm{M}(R^s)} \tag{15}$$

This shows that, given a fixed response rate $\mathrm{M}(R^s) = \mathrm{P}(R^s = 1)$, the nonresponse bias is positively related to the covariance of R^s and Y^s. Therefore, in order to hopefully reduce the nonresponse bias with the help of auxiliary variables, one should find such variables which correlate with Y^s and thereby reduce the conditional covariance of R^s and Y^s. This is illustrated by the example: $\mathrm{Cov}(Y^s, H_1^s) = 0.15$, $\mathrm{Cov}(Y^s, H_2^s) = 0$.

Notice that (15) cannot immediately be used to suggest a positive relationship between the nonresponse bias and the nonresponse rate.[12] The covariance of R^s and Y^s does not relate in any systematic way to the nonresponse rate. It depends, of course, on Y^s, and the nonresponse bias therefore depends on the variable of interest.

2.4.3 Weights Derived from Parametric Models

Using the nonresponse adjustments weights w_ω^a requires that there is at least one respondent in each adjustment cell; this restricts to some degree the possibilities of defining such cells. As an alternative, one can use the response model to define weights:

$$w_\omega^b := 1/\Pr(\dot{R} = 1 \mid \ddot{H} = H^s(\omega)) \tag{16}$$

[12] There also is only scarce empirical support for this sometimes supposed relationship; see Groves (2006).

Of course, when using $\mathrm{P}(R^s = 1 \mid H^s = H^s(\omega))$ to estimate $\Pr(\dot{R} = 1 \mid \ddot{H} = H^s(\omega))$, the weights will be identical. However, starting from (16) opens the opportunity to derive weights from parametric forms of response models. One can use, for example, a logit model

$$\Pr(\dot{R}=1 \mid \ddot{H}=h) \approx \frac{\exp(g(h;\theta))}{1+\exp(g(h;\theta))}$$

where $g(h;\theta)$ is a link function (whose specification depends on the property space of $\ddot{H}$).[13] This allows one to estimate, for each $\omega \in \mathcal{S}$, a response probability (propensity score)

$$\hat{r}(\omega) := \frac{\exp(g(H^s(\omega);\hat{\theta}))}{1+\exp(g(H^s(\omega);\hat{\theta}))} \tag{17}$$

These values can be used in two different ways.

a) One possibility is to define individual adjustment weights, say $\hat{w}^b_\omega$, proportional to $1/\hat{r}(\omega)$ and scaled to satisfy $\Sigma_{\omega\in\mathcal{S}^r}\hat{w}^b_\omega = |\mathcal{S}|$. This approach uses response probabilities only of units who actually responded (are contained in $\mathcal{S}^r$).

b) Alternatively, one can use the response probabilities of all units in $\mathcal{S}$. A simple approach uses quantiles of the distribution of these probabilities to define adjustment cells. For example, in order to define five adjustment cells one could use quintiles of the distribution of the response probabilities $\hat{r}(\omega)$.

It seems natural to form adjustment cells consisting of units with similar propensity scores (as implied when using quantiles). The following conjecture, if true, would provide an argument:

$$\text{If } |r-r_1| < |r-r_2|\,, \text{ then } \mathrm{P}[Y^s|r_H = r] \text{ is more similar to } \mathrm{P}[Y^s|r_H = r_1] \text{ than to } \mathrm{P}[Y^s|r_H = r_2]. \tag{18}$$

[13]Many different models might be employed. However, when to be used for the construction of nonresponse adjustment weights, the goal of such models is not to find optimal predictions of responses; and it is therefore difficult to see in which sense such models could be misspecified as it is sometimes suggested in the literature (e.g., da Silva and Opsomer, 2009).

Table 2 Artificial data for a sample consisting of 32 units.

ω	Y^s	R^s	H^s	ω	Y^s	R^s	H^s	ω	Y^s	R^s	H^s
1	0	1	1	11	0	1	2	23	0	1	3
2	0	1	1	12	1	1	2	24	0	1	3
3	1	1	1	13	1	1	2	25	0	1	3
4	1	1	1	14	1	1	2	26	0	1	3
5	0	0	1	15	0	1	2	27	1	1	3
6	0	0	1	16	1	1	2	28	1	1	3
7	0	0	1	17	1	1	2	29	1	1	3
8	1	0	1	18	1	1	2	30	1	1	3
9	1	0	1	19	0	0	2	31	0	0	3
10	1	0	1	20	1	0	2	32	1	0	3
				21	1	0	2				
				22	1	0	2				

This is not generally true, however, even if the conditional independence assumption (7) holds. Consider the artificial data in Table 2. The three propensity scores and corresponding distributions of Y^s are as follows:

$$\begin{array}{lll} \text{if } H^s = 1: & r_H = 0.4, & \mathrm{P}(Y^s = 1 \mid r_H = 0.4) = 0.5 \\ \text{if } H^s = 2: & r_H = 0.67, & \mathrm{P}(Y^s = 1 \mid r_H = 0.67) = 0.75 \\ \text{if } H^s = 3: & r_H = 0.8, & \mathrm{P}(Y^s = 1 \mid r_H = 0.8) = 0.5 \end{array}$$

They clearly contradict the conjecture (18). Note that Y^s and R^s are approximately independent in this example. The example also shows that propensity score variables are not always coarser than the auxiliary variables from which they are derived.

That (18) is not true is relevant for the argumentation with propensity scores. The basic argument is: Given the independence assumption (7), conditioning on propensity scores makes observed values of $\dot{R}$ uninformative about the distribution of $\dot{Y}^s$. Formally, $\Pr[\dot{Y}^s \mid r_h = r, \dot{R} = 1] = \Pr[\dot{Y}^s \mid r_h = r]$ for the model, or $\mathrm{P}[Y^s \mid r_h = r, R^s = 1] \approx \mathrm{P}[Y^s \mid r_h = r]$ for the actual observations. These relationships are no longer true, however, when $r_H = r$ is substituted by $r_H \in [r_1, r_2]$ (an interval of propensity scores). The important point is that there is no generally valid systematic relationship between propensity scores and quantities connected with the distribution of $\dot{Y}^s$ (or Y^s).

2.4.4 Combining Adjustment and Sampling Weights

So far I have assumed that $\mathcal{S}$ is a simple random sample. The aim of nonresponse adjustment then is to find a plausible estimate of the distribution of Y^s. This must be modified if units are drawn with unequal probabilities. In order to estimate $\mathrm{P}(Y = y)$, one would then use

$$\frac{1}{|\mathcal{S}|}\sum\nolimits_{\omega\in\mathcal{S}} w^s_\omega\, I[Y^s\!=\!y](\omega)$$

where w^s_ω are design-based weights, and the aim of nonresponse adjustment should be to find a plausible estimate of this quantity.

Suitably modified, one can use the argument for simple random samples. Details depend on the sampling design. To illustrate, I consider stratified sampling, based on a stratification variable K. The sample consists of subsets, say $\mathcal{S}_{(k)}$, which are simple random samples of the corresponding population strata. Analogous to (12), using adjustment cells $\mathcal{S}_{k,h} := \{\omega \in \mathcal{S} \,|\, K^s(\omega) = k, H^s(\omega) = h\}$, the assumption of approximate conditional independence can then be written as

$$\sum\nolimits_{\omega\in\mathcal{S}_{k,h}} I[Y^s\!=\!y](\omega) \;\approx\; \sum\nolimits_{\omega\in\mathcal{S}^r_{k,h}} w^c_\omega\, I[Y^s\!=\!y](\omega)$$

using adjustment weights $w^c_\omega := 1/\mathrm{P}(R^s\!=\!1 \,|\, K^s\!=\!K^s(\omega), H^s\!=\!H^s(\omega))$. It follows that

$$\sum\nolimits_{\omega\in\mathcal{S}_{(k)}} I[Y^s\!=\!y](\omega) \;\approx\; \sum\nolimits_{\omega\in\mathcal{S}^r_{(k)}} w^c_\omega\, I[Y^s\!=\!y](\omega)$$

and finally

$$\frac{1}{|\mathcal{S}|}\sum\nolimits_{\omega\in\mathcal{S}} w^s_\omega\, I[Y^s\!=\!y](\omega) \;\approx\; \frac{1}{|\mathcal{S}|}\sum\nolimits_{\omega\in\mathcal{S}^r} w^s_\omega w^c_\omega\, I[Y^s\!=\!y](\omega)$$

showing that nonresponse adjustment weights and design-based sampling weights can be combined. Notice, however, that the adjustments weights must be properly defined for the actual sampling design. If the response

proportions (or probabilities) depend on values of the stratification variable, one cannot simply use the weights w^a_ω (or w^b_ω) which are based on assuming a simple random sample.

2.5 The Quasi-Randomization Approach

It might seem tempting to draw an analogy between design-based and nonresponse adjustment weights.[14] The analogy is, however, superficial. In contrast to selection probabilities defined by a sampling design, response probabilities posited by a response model do not correspond to a random generator. Justification of nonresponse adjustment weights cannot, therefore, be based on a randomization procedure. If such weights contribute to providing better estimates, this is due to the assumption of independence between $\dot{Y}^s$ (or Y^s) and $\dot{R}$ being approximately valid in each adjustment cell.

It also follows that randomization-based arguments for the unbiasedness of an estimator cannot easily be extended to situations involving unit nonresponse. Such arguments for showing that an estimator, e.g. for $\mathrm{P}(Y\!=\!y)$, is unbiased require that one has information about values of Y for all units in the selected sample. In order to extend such arguments to situations involving unit nonresponse, one first of all would need to devise a random mechanism for the generation of realized samples. This has been called a 'quasi-randomization approach' for dealing with unit nonresponse.[15] To illustrate, one might use the response model (2) to construct

[14]For example, Little and Vartivarian (2005: 161) remarked: "Nonresponse weighting is primarily viewed as a device for reducing bias from unit nonresponse. This role of weighting is analogous to the role of sampling weights, and is related to the design unbiasedness property of the Horvitz-Thompson estimator of the total (Horvitz and Thompson 1952), which weights units by the inverse of their selection probabilities. Nonresponse weighting can be viewed as a natural extension of this idea, where included units are weighted by the inverse of their inclusion probabilities, estimated as the product of the probability of selection and the probability of response given selection; the inverse of the latter probability is the nonresponse weight."

[15]Oh and Scheuren (1983). See also Särndal, Swensson and Wretman (1992: ch. 15) where a similar approach is used.

a random generator for realized samples in the following way:

$$\text{Given } \mathcal{S}, \text{ for each } \omega \in \mathcal{S} \text{ randomly generate a value of } \dot{R}, \text{ conditional on } \ddot{H} = H^s(\omega), \text{ and define } \mathcal{S}^r \text{ as the set of units having } \dot{R} = 1. \tag{19}$$

This would allow one to define inclusion variables that are random variables w.r.t. (19): $\dot{I}^r_\omega(\mathcal{S}^r) = 1$ if $\omega \in \mathcal{S}^r$, and $= 0$ otherwise; and to derive inclusion probabilities

$$\Pr(\dot{I}^r_\omega = 1) = \sum\nolimits_{\mathcal{S}^r} \dot{I}^r_\omega(\mathcal{S}^r) \Pr(\mathcal{S}^r|H^s) = \Pr(\dot{R}=1 \,|\, \ddot{H}=H^s(\omega))$$

These variables could then be used to define an estimator for $\mathrm{P}(Y^s=y)$:[16]

$$e_{y,\mathcal{S}}(\mathcal{S}^r) := \frac{1}{|\mathcal{S}|} \sum\nolimits_{\omega\in\mathcal{S}} w^b_\omega \, I[Y^s=y](\omega) \, \dot{I}^r_\omega(\mathcal{S}^r)$$

employing the weights defined in (16). The expectation w.r.t. (19) is

$$\begin{aligned} \mathrm{E}(e_{y,\mathcal{S}}) &= \sum\nolimits_{\mathcal{S}^r} e_{y,\mathcal{S}}(\mathcal{S}^r) \Pr(\mathcal{S}^r|H^s) \\ &= \sum\nolimits_{\mathcal{S}^r} \frac{1}{|\mathcal{S}|} \sum\nolimits_{\omega\in\mathcal{S}} w^b_\omega \, I[Y^s=y](\omega) \, \dot{I}^r_\omega(\mathcal{S}^r) \Pr(\mathcal{S}^r|H^s) \\ &= \frac{1}{|\mathcal{S}|} \sum\nolimits_{\omega\in\mathcal{S}} w^b_\omega \, I[Y^s=y](\omega) \sum\nolimits_{\mathcal{S}^r} \dot{I}^r_\omega(\mathcal{S}^r) \Pr(\mathcal{S}^r|H^s) \\ &= \frac{1}{|\mathcal{S}|} \sum\nolimits_{\omega\in\mathcal{S}} w^b_\omega \, I[Y^s=y](\omega) \Pr(\dot{R}=1 \,|\, \ddot{H}=H^s(\omega)) \end{aligned}$$

Obviously, the estimator would be unbiased w.r.t. the random generator devised in (19).

However, since this random generator is purely fictitious, it is questionable whether the approach is useful. Consider the critical assumption that responses and variables of interest are (approximately) independent conditional on values of the regressor variables of the response model. This assumption would be entailed if the random generator devised in (19) really generated the realized sample. Since this is not the case, also the quasi-randomization approach must *presuppose* this assumption. The fictitious random generator therefore does not contribute any relevant argument to showing that using the adjustment weights w^b_ω (derived from the response model) provides plausible estimates.

[16]Here I assume again that $\mathcal{S}$ is a simple random sample.

3 Consideration of Functional Models

I now consider functional models that serve to formulate rules for generic units (or situations). Such models can be conceptualized either as deterministic or as probabilistic models (Rohwer 2010a, 2010b). Here I only consider probabilistic functional models (subsequently I drop the adjective 'probabilistic' and simply speak of functional models).

The most simple functional model assumes that the probability distribution of an endogenous variable, say $\dot{Y}$ (with property space $\mathcal{Y}$) depends on values of an exogenous variable, say $\ddot{X}$ (with property space $\mathcal{X}$); graphically depicted:

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \tag{20}$$

The exogenous variable $\ddot{X}$ serves to specify conditions. Since its values can be arbitrarily fixed, it can be conceived of neither as a statistical nor as a random variable. To remind of its special status as an exogenous variable without an associated distribution it is marked by two dots. Since $\ddot{X}$ has no distribution, there also is no distribution for $\dot{Y}$ (and it is therefore not a random variable in the usual sense of the word). However, in order to make quantitative statements possible, one can think of distributions of $\dot{Y}$ if particular values of $\ddot{X}$ are fixed. To make this idea explicit, one uses a stochastic function, $x \longrightarrow \Pr[\dot{Y} \,|\, \ddot{X}=x]$, that assigns to each value x of $\ddot{X}$ a probability distribution of the variable $\dot{Y}$.

Functional models can be estimated with data from random samples.[17] Details depend on whether and how the models are parameterized. Here I only consider difficulties which could result from unit nonresponse. The basic idea is to integrate a response model into the primarily interesting functional model and then consider reduced models resulting from conditioning on response ($\dot{R}=1$).

[17]It would be possible to think of the functional model as intending a description of the population from which the sample is drawn. However, the general notion of a functional model does not require its linkage to any particular target population.

3.1 Basic Forms of Combined Models

I assume that (20) is the model of primary interest and there is a further endogenous variable, $\dot{R}$, indicating how the sampled data relate to the model. The data only allow one to estimate $\Pr[\dot{Y} \mid \ddot{X}=x, \dot{R}=1]$, and the question is whether, and how, one can estimate $\Pr[\dot{Y} \mid \ddot{X}=x]$.

(A) A first situation occurs when the response variable only depends on exogenous variables:

$$\begin{tikzcd} \ddot{X} \arrow[r, ->>] \arrow[d, ->>] & \dot{Y} \\ \dot{R} & \end{tikzcd} \tag{21}$$

This entails that, conditional on values of $\ddot{X}$, $\dot{Y}$ and $\dot{R}$ are stochastically independent:

$$\Pr(\dot{Y}=y, \dot{R}=1 \mid \ddot{X}=x) \;=\; \Pr(\dot{Y}=y \mid \ddot{X}=x)\,\Pr(\dot{R}=1 \mid \ddot{X}=x) \tag{22}$$

and consequently

$$\Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{R}=1) \;=\; \Pr(\dot{Y}=y \mid \ddot{X}=x) \tag{23}$$

In order to estimate the model of interest, one can use the data from the realized sample $\mathcal{S}^r$ without the need to adjust for unit nonresponse. Of course, a consequence of nonresponse could be that for some values (or regions) in the property space of $\ddot{X}$ no, or only very few, observations are available, and this must then be taken into account.

(B) Equation (22) will be valid if $\dot{R}$, but not $\dot{Y}$, also depends on values of a further exogenous variable, say $\ddot{Z}$. A possibly different situation occurs when also $\dot{Y}$ depends on $\ddot{Z}$:

$$\begin{tikzcd} \ddot{X} \arrow[r, ->>] \arrow[d, ->>] & \dot{Y} \\ \dot{R} & \ddot{Z} \arrow[l, ->>] \arrow[u, ->>] \end{tikzcd} \tag{24}$$

This formally equals (21) with an exogenous variable $(\ddot{X}, \ddot{Z})$ consisting

of two components. A new situation occurs, however, when values of $\ddot{Z}$ cannot be observed. First of all, it is necessary then to explicitly define the reduced model that one intends to estimate. Note that already the definition of this reduced model requires either to fix a specific value of $\ddot{Z}$, or to substitute $\ddot{Z}$ by a variable for which one can assume a distribution (see Rohwer 2010a: 52ff).

Assume that $\ddot{Z}$ is substituted by a random variable, $\dot{Z}$, with an unknown distribution, but the structure of the model is not changed so that $\dot{Z}$ is still an exogenous variable and independent of $\ddot{X}$. One might then be interested in a reduced model, $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$, defined by

$$\Pr(\dot{Y}=y \mid \ddot{X}=x) = \sum\nolimits_z \Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{Z}=z)\,\Pr(\dot{Z}=z)$$

Additional conditioning on $\dot{R}=1$ results in

$$\begin{aligned} &\Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{R}=1) = \\ &\qquad \sum\nolimits_z \Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{Z}=z)\,\Pr(\dot{Z}=z \mid \ddot{X}=x, \dot{R}=1) \end{aligned}$$

Since $\Pr(\dot{Z}=z \mid \ddot{X}=x, \dot{R}=1) \neq \Pr(\dot{Z}=z)$, unit nonresponse cannot be ignored; and since $\dot{Z}$ cannot be observed, there is no way to eliminate the nonresponse error.

(C) A further situation occurs when the response variable also depends on values of an endogenous variable in the model of primary interest:

$$\begin{tikzcd} \ddot{X} \arrow[r, ->>] \arrow[d, ->>] & \dot{Y} \arrow[dl, ->>] \\ \dot{R} & \end{tikzcd} \tag{25}$$

The conditional independence formulated in (22) is no longer valid. The available data only allow one to estimate $\Pr[\dot{Y}|\ddot{X}=x, \dot{R}=1]$, and there are no possibilities to empirically assess deviations from $\Pr[\dot{Y}|\ddot{X}=x]$.

It would not even suffice to know, or make assumptions about, the response model $(x, y) \longrightarrow \Pr(\dot{R}=1 \mid \ddot{X}=x, \dot{Y}=y)$. The relationship between the model of interest and the model that can be estimated with the realized sample is given by

$$\Pr(\dot{Y}=y \mid \ddot{X}=x) = \Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{R}=1)\,\frac{\Pr(\dot{R}=1 \mid \ddot{X}=x)}{\Pr(\dot{R}=1 \mid \ddot{X}=x, \dot{Y}=y)} \tag{26}$$

showing that the joint distribution of $\dot{Y}$ and $\dot{R}$, conditional on values of $\ddot{X}$, would be required.

3.2 Bias Reduction with Auxiliary Variables

Mainly two strategies have been proposed for situations in which responses depend on an endogenous variable of a functional model. One strategy relies on specific assumptions about the mathematical form of the joint distribution of $\dot{Y}$ and $\dot{R}$. Another strategy that will be considered in the present section is similar to the use of nonresponse adjustment weights in the context of descriptive estimation.

I refer to the combined model (25). One is interested in $\Pr[\dot{Y}|\ddot{X}=x]$, but the data only allow one to estimate $\Pr[\dot{Y}|\ddot{X}=x, \dot{R}=1]$. In order to assess the nonresponse error, one may use the equation

$$\mathrm{Cov}(\dot{Y},\dot{R} \mid \ddot{X}=x) = \mathrm{E}(\dot{Y} \mid \ddot{X}=x, \dot{R}=1)\,\mathrm{E}(\dot{R} \mid \ddot{X}=x) - \mathrm{E}(\dot{Y} \mid \ddot{X}=x)\,\mathrm{E}(\dot{R} \mid \ddot{X}=x)$$

This is similar to (14), but now conditional on values of $\ddot{X}$, and implies

$$\mathrm{E}(\dot{Y} \mid \ddot{X}=x, \dot{R}=1) - \mathrm{E}(\dot{Y} \mid \ddot{X}=x) = \frac{\mathrm{Cov}(\dot{Y},\dot{R} \mid \ddot{X}=x)}{\mathrm{E}(\dot{R} \mid \ddot{X}=x)} \tag{27}$$

showing that the nonresponse error is positively related to the covariance between $\dot{R}$ and $\dot{Y}$, conditional on values of $\ddot{X}$. One should therefore try to find auxiliary variables such that additional conditioning on these variables diminishes the correlation between $\dot{R}$ and $\dot{Y}$.

Instead of (25), there is then an enlarged model

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \arrow[d] \arrow[dl] \\ \dot{R} & \dot{H} \arrow[l] \end{tikzcd} \tag{28}$$

which includes an endogenous auxiliary variable $\dot{H}$. If $\dot{H}$ adds to the prediction of $\dot{R}$, it is likely that additional conditioning on this variable will reduce (albeit in an unknown amount) the nonresponse error due to the correlation between $\dot{Y}$ and $\dot{R}$.

Sometimes it might be possible to find auxiliary variables such that the direct arrow from $\dot{Y}$ to $\dot{R}$ can be omitted:

$$\begin{tikzcd} \ddot{X} \arrow[r] \arrow[d] & \dot{Y} \arrow[d] \\ \dot{R} & \dot{H} \arrow[l] \end{tikzcd} \tag{29}$$

If this is justified, the model entails the conditional independence relation

$$\Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{H}=h, \dot{R}=1) = \Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{H}=h)$$

allowing to derive

$$\Pr(\dot{Y}=y \mid \ddot{X}=x) = \sum\nolimits_h \Pr(\dot{Y}=y \mid \ddot{X}=x, \dot{H}=h, \dot{R}=1)\,\Pr(\dot{H}=h \mid \ddot{X}=x)$$

However, in order to recover $\Pr(\dot{Y}=y|\ddot{X}=x)$, one would need the conditional probabilities $\Pr(\dot{H}=h \mid \ddot{X}=x)$ which cannot be estimated without bias when observations are conditional on $\dot{R}=1$, even if the auxiliary variable could be observed unconditionally.

3.3 Digression on Propensity Scores

Propensity scores, as introduced in section 2.3.3, concern responses of units selected for inclusion in a sample. In this context, propensity scores are identical with response probabilities modeled as being dependent on auxiliary variables in such a way that, conditional on these variables, responses and variables of interest are (hopefully) approximately independent. In this section I briefly compare this understanding with a usage of propensity scores, suggested by Rosenbaum and Rubin (1983), that intends to simulate randomization w.r.t. confounders in observational studies of causal effects.

In this context, $\dot{R}$ represents not a response, but the presence of a causal factor. Instead of (25), the functional model is then given by

$$\begin{tikzcd} \ddot{X} \arrow[r, twoheadrightarrow] \arrow[d] & \dot{Y} \\ \dot{R} \arrow[ur, twoheadrightarrow] & \end{tikzcd} \tag{30}$$

The outcome variable, $\dot{Y}$, depends on values of $\dot{R}$ and $\ddot{X}$. $\dot{R}$ represents the presence ($\dot{R}=1$) or absence ($\dot{R}=0$) of a particular causal factor; $\ddot{X}$ represents additional conditions on which $\dot{Y}$ depends. Hinted at by calling $\ddot{X}$ a 'confounder', the interest concerns a reduced model, $\dot{R} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$, where $\dot{Y}$ depends only on the particular causal factor. The causal effect of this factor could then be defined by

$$\Pr[\dot{Y} \mid \dot{R}=1] - \Pr[\dot{Y} \mid \dot{R}=0] \tag{31}$$

However, this reduced model cannot be derived from (30). In other words, given this model, a causal effect as presupposed in (31) does not exist. In order to define this effect, one would need a completely different model, namely

$$\begin{tikzcd} \dot{X} \arrow[r, twoheadrightarrow] & \dot{Y} \\ \ddot{R} \arrow[u] \arrow[ur, twoheadrightarrow] & \end{tikzcd} \tag{32}$$

where the confounding variable is endogenous and the treatment variable is exogenous. This model would allow one to derive the reduced model which is presupposed by (31), namely

$$\Pr[\dot{Y} \mid \ddot{R}=r] = \sum\nolimits_x \Pr[\dot{Y} \mid \dot{X}=x, \ddot{R}=r] \Pr(\dot{X}=x \mid \ddot{R}=r) \tag{33}$$

The idea of randomization w.r.t. the confounding variable starts from the model (32). The idea is to generate a situation in which $\Pr[\dot{X} \mid \ddot{R}=1] = \Pr[\dot{X} \mid \ddot{R}=0]$. But this requires that the arrow from $\ddot{R}$ to $\dot{X}$ can be dropped and $\dot{X}$ can be changed into an exogenous variable.

This might be possible in an experimental context where values of $\ddot{R}$ can be randomly assigned to units. However, this cannot be done in observational studies when treatment variables are endogenous as assumed in the model (30). This model requires to take seriously that the outcome variable depends on values of $\dot{R}$ and $\ddot{X}$ and that these variables are not independent. Of course, the model entails the stochastic function

$$(r,x) \longrightarrow \Pr[\dot{Y} \mid \dot{R}=r, \ddot{X}=x] \tag{34}$$

which can be used to define conditional effects:

$$\Pr[\dot{Y} \mid \dot{R}=1, \ddot{X}=x] - \Pr[\dot{Y} \mid \dot{R}=0, \ddot{X}=x] \tag{35}$$

It is now easy to understand what can be achieved with propensity scores for 'treatment assignment'. These are probabilities $\Pr(\dot{R}=1|\ddot{X}=x)$, taken as values of a variable, say $\ddot{S}$.[18] Explicitly defined:

$$\ddot{S}=s \iff \ddot{X} \in \mathcal{X}_s := \{x \in \mathcal{X} \mid \Pr(\dot{R}=1 \mid \ddot{X}=x) = s\}$$

This variable can be used instead of $\ddot{X}$ in (34), but the only gain is a possibly coarser formulation of the dependency on values of $\ddot{X}$:

$$(r,s) \longrightarrow \Pr[\dot{Y} \mid \dot{R}=r, \ddot{S}=s] = \Pr[\dot{Y} \mid \dot{R}=r, \ddot{X} \in \mathcal{X}_s] \tag{36}$$

As stressed by Rosenbaum and Rubin (1983), $\dot{R}$ is independent of $\ddot{X}$, conditional on values of $\ddot{S}$; but this fact does not lead to a kind of randomization. One still can only define conditional effects (by substituting $\ddot{X}=x$ with $\ddot{S}=s$ in (35)).

This finally reveals a further difference between the two uses of propensity scores. When constructing propensity scores for responses in surveys it might well be possible to find auxiliary variables such that, conditional

[18] $\ddot{S}$ is an exogenous variable because it is derived from the exogenous variable $\ddot{X}$.

on their values, responses and variables of interest become approximately independent. In contrast, starting from the model (30), propensity scores for 'treatment assignment' ($\dot{R} = 1$) are already fixed by the model and therefore cannot be freely constructed in such a way that these assignments become independent of further causally relevant conditions.

4 Conclusion

The main conclusion of the foregoing discussion is that there is no generally applicable method for successfully coping with unit nonresponse. Each usage of data from samples with a relevant proportion of unit nonresponse requires a separate consideration of whether the nonresponses can be ignored or, if not, whether and how nonresponse bias can be reduced. This conclusion leads to two suggestions for the arrangement of data sets:

a) Data sets should not be supplemented with ready-made nonresponse adjustment weights.

b) Instead, the data set should be supplemented with a list of all units originally selected, and values of all variables available for the complete sample should be added.

References

da Silva, D. N., Opsomer, J. D. (2009). Nonparametric Propensity Weighting for Survey Nonresponse Through Local Polynomial Regression. *Survey Methodology* 35, 165–176.

Deville, J.-C., Särndal, C.-E., Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association* 88, 1013–20.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70, 646–675.

Holt, D., Elliot, D. (1991). Methods of Weighting for Unit Non-Response. *The Statistician* 40, 333–342.

Little, R. J., Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology* 31, 161–168.

Oh, H. L., Scheuren, F. J. (1983). Weighting Adjustment for Unit Nonresponse. In: W. G. Madow, I. Olkin, D. B. Rubin (eds.), *Incomplete Data in Sample Surveys*, Vol. 2, 143–184. New York: Academic Press.

Rohwer, G. (2010a). *Models in Statistical Social Research*. London: Routledge.

Rohwer, G. (2010b). Qualitative Comparative Analysis. A Discussion of Interpretations. *European Sociological Review* [published online July 19, 2010].

Rosenbaum, P. R., Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 41–55.

Särndal, C.-E., Swensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.