Correcting for Survey Misreports using Auxiliary Information with an Application to Estimating Turnout

Jonathan N. Katz
California Institute of Technology

Gabriel Katz
California Institute of Technology

August 8, 2008

ABSTRACT

Misreporting is a problem that plagues researchers that use survey data. In this paper, we develop a parametric model that corrects for misclassified binary responses using information on the misreporting patterns obtained from auxiliary data sources. The model is implemented within the Bayesian framework *via* Markov Chain Monte Carlo (MCMC) methods, and can be easily extended to address other problems exhibited by survey data, such as missing response and/or covariate values. Using Monte Carlo simulations, we show that, even with small rates of misclassification, our proposed solution improves estimates and inference with respect to standard models ignoring misreporting, and it also outperforms other methods proposed in the literature when misreporting is associated with the covariates affecting the true response variable. While the model is fully general, we illustrate its application in the context of estimating models of turnout using data from the American National Elections Studies.

Much of the empirical work in the social sciences is based on the analysis of survey data. However, as has been widely documented (Battistin 2003; Bound, Brown and Mathiowetz 2001; Poterba and Summers 1986), these data are often plagued by measurement errors. There are many possible sources for such errors. Interviewers may erroneously record answers to survey items, and respondents may provide inaccurate responses due to an honest mistake, misunderstanding or imperfect recall (Gems, Ghaosh and Hitlin 1982; Hausman, Abrevaya and Scott-Morton 1998; Molinari 2003). Also, as underscored by the social psychology literature, survey respondents tend to overreport socially desirable behaviors and underreport socially undesirable ones (Cahalan 1968; Loftus 1975). In the case of discrete or categorical variables, mismeasurement problems have been traditionally referred to as "misclassification" errors (Aigner 1973; Bollinger 1996; Bross 1954).

In the political science literature, concerns about misclassification have been particularly prevalent in the analysis of voting behavior. Empirical studies of the determinants of voter turnout focus on how the probability of an individual voting varies according to relevant observable factors, such as citizen's level of political information, registration laws, or demographic characteristics. That is, these studies are interested in estimating the conditional distribution of the turnout decision given certain characteristics of interest.[1] The decision to vote, however, is typically not observed due to the use of secret ballot in the U.S. Furthermore, even if we could observe turnout from the official

[1]The literature is far too vast to even begin to fully cite here. See Aldrich (1993) for a review of the theoretical literature and Wolfinger and Rosenstone (1980) for an influential empirical study.

ballots we would not, in general, be able to observe all the characteristics — e.g., the voter's policy preferences or information about the candidates — that presumably affect the decision. Hence, political scientist rely on the use of survey instruments, such as the American National Election Study (ANES) or the Current Population Survey (CPS), that include both measures of respondents' relevant characteristics and their *self-reported* voting behavior. This almost always leads to estimation of the common logit or probit models, since the turnout decision is dichotomous, although there are alternatives such as scobit (Nagler 1994) or non-parametric models (Härdle 1990) for discrete choice models.

However, it has been long established that some survey respondents misreport voting, i.e., they report that they have voted when in fact they did not do so (Burden 2000; Clausen 1968; Katosh and Traugott 1981; Miller 1952; Parry and Crossley 1950; Sigelman 1982; Silver, Anderson and Abramson 1986). The evidence that misreporting is a problem can be found in a series of validation studies that the ANES conducted in 1964, 1976, 1978, 1980, 1984, 1988 and 1990. These validation studies were possible, but expensive, because voting is a matter of public record, although for whom a voter voted is not. After administering a post-election survey to a respondent, an official from the ANES was sent to the respondent's local registrar of elections to see if in fact they were recorded as having voted in the election. This is not an easy task, since respondents often do not know where they voted, election officials differ in their ability to produce the records in a usable form, and there might be differences between the survey data and the public records due to errors in spelling or recording. This means that the validated data may also be mismeasured, but for this paper we will assume it is correct. That said, the ANES for these years included both the respondent's *self-reported* vote and the *validated* vote. The differences between the two measures are fairly shocking. Depending on the election year, between 13.6 and 24.6 percent of the respondents claiming to have voted did in fact not according to the public records.[2] In contrast, only between 0.6% and 4.0% of the respondents in the 1964 – 1990 validated surveys who reported not having voted did vote according to the official records. Since there is no reason to believe that measurement errors should mainly be of false positives — i.e., reporting voting when the official record contradicts this claim —, this lends some credence to the social pressures argument for misreporting (Bernstein, Chadha and Montjoy 2001) and should help mitigate some of our concerns about other potential sources of classification errors, such as inaccurate records.[3] The large differences between reported and validated turnout led to a cottage industry analyzing the causes of misreporting (Abramson and Claggett 1984, 1986*a*,*b*, 1991; Ansolabehere and Hersh 2008; Cassel 2003; Hill and Hurley 1984; Katosh and Traugott 1981; Sigelman 1982; Silver, Anderson and Abramson 1986; Weir 1975) and to a debate about how to best measure misreporting (Anderson and Silver 1986). All of these studies find that misreporting varies systematically with some characteristics of interest, but none of them provides an estimation solution to correct for possible misreporting. The open question then is what to do about the problem of respondents misreporting. One possibility would be to use only validated data. At some level this is an appealing option. If we are sure that the validated data is correct, then estimation and inference is straightforward. Unfortunately, collecting the validated turnout data is difficult and expensive, and ANES has stopped doing validation studies for these reasons. Furthermore, even if validation studies were free, some states, such as Indiana, make it impossible to validate votes. Hence, if we are going to limit ourselves to use only fully

[2]The Current Population Survey (CPS) also exhibits considerable turnout overreporting, although the magnitude is substantially lower than for the ANES (Highton 2004). As shown by Hausman, Abrevaya and Scott-Morton (1998) and Neuhaus (1999), however, even modest amounts of misreporting can affect parameter estimates.

[3]Clearly, other reasons besides social desirability may also contribute to explain differences between self-reported and validated turnout (Abelson, Loftus and Greenwald 1992).

validated data, our samples will be much smaller. Moreover, would also be throwing away the useful information included in the already collected but non-validated studies.[4] On the other hand, simply ignoring misreporting and using self-reported turnout to estimate standard probit or logit models can result in biased and inconsistent parameter estimates and inaccurate standard errors, potentially distorting the relative impact of the characteristics of interest on the response variable and leading to erroneous conclusions (Davidov, Faraggi and Reiser 2003; Hausman, Abrevaya and Scott-Morton 1998; Neuhaus 1999).[5]

In this paper we develop a simple Bayesian approach to correct for misreporting, allowing researchers to continue to use the self-reported data while improving the accuracy of the estimates and inferences drawn in the presence of misclassified binary responses.[6] Our model draws on Hausman, Abrevaya and Scott-Morton (1998), but incorporates information on the misreporting process from auxiliary data sources, aiding in identification (Gu 2006; Molinari 2003) and making it easier to avoid the problems that limit the use of Hausman, Abrevaya and Scott-Morton (1998)'s modified maximum likelihood estimator in small samples such as those typically used in political science (Christin and Hug 2004; Gu 2006). While incorporating this information into the analysis of the sample of interest using frequentist methods is far from straightforward (Prescott and Garthwaite 2005), this can be easily accomplished within the Bayesian framework *via* Markov Chain Monte Carlo (MCMC) simulations. Although other Bayesian approaches have been proposed to adjust for misclassification using prior information to overcome fragile or poor identifiability, they either rely exclusively on elicitation of experts' opinions (McInturff et al. 2004; Paulino, Soares and Neuhaus 2003) or assume that information on both the true and the fallible response is available for all subjects in a random subsample of the data (Viana 1994; Prescott and Garthwaite 2002, 2005). In contrast, the information on the misreport patterns incorporated into our model need not come from the sample of interest, and can be combined with elicitation of experts' beliefs if needed. In the empirical application presented in this paper we will use earlier and small-sample validation studies to correct for misreporting. However, matched official records, administrative registers and possibly even aggregate data might be used to gain this information. Given the potential difficulties of eliciting probabilities from experts' opinions and the scarcity of internal validation designs relative to administrative data sets, external validation studies and other sources of ancillary information (Bound, Brown and Mathiowetz 2001; Garthwaite, Kadane and OHagan 2004; Hu and Riddert 2007; Wiegmann 2005), the correction developed in this paper provides a more flexible way of incorporating prior information and can be more widely applied than existing approaches.[7] In addition, these alternative approaches focus only on the case in which the misclassification rates are independent of all covariates. As mentioned above, this assumption seems to be inappropriate in the case of the determinants of voter turnout, as well as in many other potential applications. The

[4] In the case of the ANES, turnout is one of the few survey items included since the late 1940s, covering a larger period than any other continuing survey (Burden 2000). Validation studies, on the other hand, only comprise a handful of elections.

[5] A third strand of research focuses on procedures for reducing the frequency of overreporting, such as altering question wording or reformulating survey questions (Belli, Traugott and Rosenstone 1994; Belli et al. 1999; Bound, Brown and Mathiowetz 2001). Nonetheless, while this can improve the quality of future datasets, we would still be wasting large amounts of data collected in previous surveys.

[6] We focus on the case of misclassified responses and error-free covariates. Several methods have been proposed to adjust for measurement error in the covariates. See Carroll, Ruppert and Stefanski (1995) and Thürigen et al. (2000) for a review.

[7] In internal validation studies, the true response is available for a subset of the main study and can be compared to the imperfect or observed response. In the case of the external validation designs, the misreport pattern is estimated using data outside the main study.

magnitude and direction of the biases when misreporting is covariate-dependent can be quite different than in the case of constant misclassification rates (Davidov, Faraggi and Reiser 2003; Neuhaus 1999) and, in the context of analyzing voting behavior, Bernstein, Chadha and Montjoy (2001) show that ignoring the correlation between the covariates of interest and the misreport probabilities may seriously distort multivariate explanations of the turnout decision. Finally, our approach enables us to simultaneously address another important problem with survey data, namely missing outcome and/or covariate values, using fully Bayesian model-based imputation (Ibrahim et al. 2005).

Although our model is developed in the context of estimating the conditional probability of turning out to vote, the method is general and will be applicable whenever misclassification of a binary outcome in a survey is anticipated and there is auxiliary information on the misreporting patterns. For instance, our approach could be used to analyze survey data on participation in social welfare programs (Hernanz, Malherbet and Pellizzari 2004), pension plans (Molinari 2003), energy consumption (Gu 2006), employment status (Hausman, Abrevaya and Scott-Morton 1998) and many other areas where we expect to see substantial rates of misreporting and potential correlation between some of the covariates affecting the response and the misreport probabilities. The model can also be implemented when misreporting depends on covariates other than those influencing the outcome. For example, for a substantial proportion of the CPS sample, turnout is measured by proxy, rather than self-reported (Highton 2004). In this case, the misclassification probabilities would be modeled using information on misreporting patterns among household members reporting other members' turnout decision, which could be obtained from validated CPS studies.[8] Extensions of our method to discrete choice models with more than two categories along the lines of Abrevaya and Hausman (1999) and Dustmant and van Soest (2004) are possible as well.

The paper proceeds as follows. The next section formally lays out the estimation problem in the presence of misreporting and develops our proposed solution. Section 2 presents results from a Monte Carlo experiment illustrating how important misreporting can be in practice and comparing the estimates from our method with those obtained under several alternative approaches. We also evaluate the robustness of our approach to misspecification of the misreport model and assess its performances in the presence of both misclassification and missing data. In Section 3, we provide three applications of our methodology using data on voter turnout from the ANES. Finally, Section 4 concludes.

1. CORRECTING FOR MISREPORTING IN BINARY CHOICE MODELS

1.1. *Defining the Problem*

Let y_i be a dichotomous (dummy) variable, and denote by $\mathbf{x}_i$ a vector of individual characteristics of interest. We want to estimate the conditional distribution of y_i given $\mathbf{x}_i$, $\Pr[y_i|\mathbf{x}_i]$. However, instead of observing the "true" dependent variable y_i, assume we observe the self-reported indicator $\tilde{y}_i$. Most studies use the observed $\tilde{y}_i$ as the dependent variable, typically running either a probit or logit model to estimate $\Pr[\tilde{y}_i = 1|\mathbf{x}_i]$.

In order to know whether this substitution can lead to incorrect inferences, we need to know the relationship between $\Pr[\tilde{y}_i = 1|\mathbf{x}_i]$ and $\Pr[y_i = 1|\mathbf{x}_i]$. We can always write

$$\begin{aligned}\Pr[\tilde{y}_i = 1|\mathbf{x}_i] = \Pr[\tilde{y}_i = 1|\mathbf{x}_i, y_i = 1] \cdot \Pr[y_i = 1|\mathbf{x}_i] + \\ \Pr[\tilde{y}_i = 1|\mathbf{x}_i, y_i = 0] \cdot \Pr[y_i = 0|\mathbf{x}_i],\end{aligned} \tag{1}$$

[8] We thank an anonymous referee for pointing us to this potential application of our model.

by the law of total probability. All that we have done is to rewrite the probability $\Pr[\tilde{y}_i = 1|\mathbf{x}_i]$ into two components: when the self-reported or observed variable $\tilde{y}_i$ coincides with the true response y_i, and when it does not. Also, noting that $\Pr[\tilde{y}_i = 0|\mathbf{x}_i, y_i = 1] = 1 - \Pr[\tilde{y}_i = 1|\mathbf{x}_i, y_i = 1]$ we can re-write the relationship as

$$\Pr[\tilde{y}_i = 1|\mathbf{x}_i] = (1 - \pi_i^{1|0} - \pi_i^{0|1}) \Pr[y_i = 1|\mathbf{x}_i] + \pi_i^{1|0} \tag{2}$$

where $\pi_i^{1|0} = \Pr[\tilde{y}_i = 1|y_i = 0, \mathbf{x}_i]$ is the probability that the respondent falsely claims $\tilde{y}_i = 1$ when in fact $y_i = 0$, and $\pi_i^{0|1} = \Pr[\tilde{y}_i = 0|y_i = 1, \mathbf{x}_i]$ is the probability the observed response takes the value 0 when the true response is $y_i = 1$. It is important to note that the probability of each type of misreporting is conditional on $\mathbf{x}_i$.

Standard methods for estimating binary choice models generally assume that the conditional distribution of the dependent variable given $\mathbf{x}_i$ is known up to a parameter vector β. However, unless $\pi_i^{0|1} = \pi_i^{1|0} = 0 \quad \forall i$, estimating the conditional probability $\Pr[\tilde{y}_i = 1|\mathbf{x}_i]$ rather than $\Pr[y_i = 1|\mathbf{x}_i]$ will generally lead to biased estimates of β and inaccurate standard errors, with even small probabilities of misreporting potentially leading to significant amounts of bias (Davidov, Faraggi and Reiser 2003; Hausman, Abrevaya and Scott-Morton 1998; Neuhaus 1999). In addition, the marginal effect of covariate x on the observed response $\tilde{y}_i$ and on the true response y_i will differ by

$$\begin{aligned} \frac{\partial \Pr[\tilde{y}_i = 1|\mathbf{x}_i]}{\partial x} - \frac{\partial \Pr[y_i = 1|\mathbf{x}_i]}{\partial x} = &- \left(\frac{\partial \pi_i^{1|0}}{\partial x} + \frac{\partial \pi_i^{0|1}}{\partial x} \right) \Pr[y_i = 1|\mathbf{x}_i] \\ &- (\pi_i^{1|0} + \pi_i^{0|1}) \frac{\partial \Pr[y_i = 1|\mathbf{x}_i]}{\partial x} + \frac{\partial \pi_i^{1|0}}{\partial x}. \end{aligned} \tag{3}$$

As a result, inferences drawn on the relationship between the covariates of interest and the response variable may change substantially when estimated based on the likelihood function defined by $\Pr[\tilde{y}_i = 1|\mathbf{x}_i]$ rather than on the true model $\Pr[y_i = 1|\mathbf{x}_i]$, depending on the distribution of $\beta'\mathbf{x}_i$ and the covariate vector $\mathbf{x}_i$, on the prevalence of misclassification and on the relationship between the probabilities of misreporting and the covariates in $\mathbf{x}_i$ (Bernstein, Chadha and Montjoy 2001; Hausman, Abrevaya and Scott-Morton 1998; Neuhaus 1999).

Different parametric models have been proposed to correct for misclassification of the dependent variable in binary choice models (Carroll, Ruppert and Stefanski 1995; Hausman, Abrevaya and Scott-Morton 1998; Marshall 1990; McInturff et al. 2004; Morrissey and Spiegelman 1999; Paulino, Soares and Neuhaus 2003; Prescott and Garthwaite 2002, 2005).[9] In particular, Hausman, Abrevaya and Scott-Morton (1998) proposed a modified maximum likelihood estimator that requires the "monotonicity" condition $\pi_i^{1|0} + \pi_i^{0|1} < 1$ to achieve identification. Using Monte Carlo simulations, they showed that their model consistently estimates the extent of misclassification and the parameter vector β, at least in large samples. More recently, however, Christin and Hug (2004) replicated the work of Hausman, Abrevaya and Scott-Morton (1998) for different sample sizes, and

[9] A comprehensive review of different methods developed to deal with misclassification and measurement errors in nonlinear models can be found in Carroll, Ruppert and Stefanski (1995).

found that the modified maximum likelihood estimator performed consistently better than simple probit models ignoring misclassification only in samples of 5,000 or more observations. In smaller samples, standard probit estimators outperformed it in many cases, and Christin and Hug (2004) concluded that the modified maximum likelihood estimator is only advisable for large samples. As noted by Gu (2006), the failure of Hausman, Abrevaya and Scott-Morton (1998)'s estimator in small samples is likely due to the insufficiency of the monotonicity condition to ensure model identification. For such sample sizes typically available in political science, even moderate rates of misclassification may hinder model identification, so different assumptions may be required to put bounds on the misclassification rates and the regression coefficients. In addition, Hausman, Abrevaya and Scott-Morton (1998) and, in fact, most empirical applications of models proposed to correct for misreporting, assume constant misclassification rates, failing to account for the potential influence of the covariates of interest on $\pi^{1|0}$ and $\pi^{0|1}$.[10] Relevant prior information on the misreport patterns is often available from auxiliary data sources, such as internal or external validation studies, small sample pilots or administrative registers, which can be used to impose restrictions on the misreport probabilities and regression coefficients to aid in identification and improve inferences on the relationship between $\mathbf{x}$ and $\mathbf{y}$ (Chen 1979; Molinari 2003).

In order to incorporate the information on the misreporting structure from auxiliary data sources, we propose a Bayesian approach based on Markov Chain Monte Carlo (MCMC) methods. This approach has three basic advantages in this setting. First, results from previous statistical studies can be easily incorporated into the model for the sample of interest within the Bayesian framework (Dunson and Tindall 2000; Ibrahim and Chen 2000; Ibrahim, Ryan and Chen 1998). Second, MCMC methods directly account for the extra uncertainty in the variances caused by using estimates of the misreport probabilities obtained from the auxiliary data instead of their true values. In contrast, in the context of frequentist estimation, this would require additional "post-estimation" steps, such as bootstrapping (Haukka 1995), applying the results of Murphy and Topel (1985) for two-step estimators, or using numerical techniques (Kuha 1994).[11] In addition, our approach does not rely on large sample assumptions and avoids the need for complicated numerical approximations (Viana 1994) when the posterior distributions are analytically intractable. The model is a simple modification of Hausman, Abrevaya and Scott-Morton (1998)'s estimator and can be easily implemented by practitioners and applied researchers using flexible and freely available software for Bayesian analysis such as WinBUGS or JAGS (Plummer 2009; Spiegelhalter, Thomas and Best 2003).

1.2. *A Bayesian Model to Correct for Misreporting using Auxiliary Data*

We are interested in accurately estimating the effect of the individual characteristics of interest on the conditional distribution of the true response. Hence, the focus of our analysis lies in the marginal posterior distribution of β, while the the modelization of the conditional probabilities $\pi_i^{1|0}$ and $\pi_i^{0|1}$ can be regarded as "instrumental".

Since the observed response variable is dichotomous, we can start by assuming that, conditional

[10] Abrevaya and Hausman (1999); Hausman, Abrevaya and Scott-Morton (1998) and Paulino, Soares and Neuhaus (2003), among others, discuss extensions to deal with covariate-dependent misclassification, but they do not analyze this case in practice.

[11] Another possible approach is to assume that misclassification rates are known and equal to those prevalent in the auxiliary data (Poterba and Summers 1995). Nonetheless, as noted by Hausman, Abrevaya and Scott-Morton (1998), not only will this lead to inconsistent parameter estimates if the assumed misclassification probabilities are not consistent estimates of the true probabilities, but the standard errors of the coefficient estimates will be understated.

on some set of relevant individual characteristics, the observations are independently and identically distributed according to a Bernoulli distribution — as in Hausman, Abrevaya and Scott-Morton (1998). The probability of the sample can therefore be written as

$$\mathcal{L}(\theta|\tilde{\mathbf{y}}, \mathbf{x}) = \prod_{i=1}^{N} \Pr[\tilde{y}_i|\mathbf{x}_i, \theta]^{\tilde{y}_i}(1 - \Pr[\tilde{y}_i|\mathbf{x}_i, \theta])^{1-\tilde{y}_i}, \tag{4}$$

with $\theta = \{\pi_i^{1|0}, \pi_i^{0|1}, \beta\,'\}$. We will further assume that the conditional probability of the true response variable is given by $\Pr[y_i = 1|\mathbf{x}_i] = F(\beta'\mathbf{x}_i)$, where $F(\cdot)$ is some cumulative density function. For ease of exposition, we use the probit link, so that $F(\cdot)$ is the standard normal distribution denoted by $\Phi(\cdot)$. This will lead to a probit model with a correction for misreport; the use of the logit link function would result in a logit model with a correction for misreporting. We also assume that $\Pr[y_i = 1|\mathbf{x}_i]$ is *a priori* independent of $\pi_i^{1|0}$ and $\pi_i^{0|1}$.[12] Substituting for $\Pr[\tilde{y}_i|\mathbf{x}_i, \theta]$ in Equation 2 and denoting by $\mathcal{S}$ the sample data, we arrive at:

$$\begin{aligned}\mathcal{L}(\beta, \pi_i^{1|0}, \pi_i^{0|1}|\mathcal{S}) = \prod_{i=1}^{N}\Bigg[&\left[(1 - \pi_i^{1|0} - \pi_i^{0|1})\Phi(\beta\,'\mathbf{x}_i) + \pi_i^{1|0}\right]^{\tilde{y}_i} \\ &\times \left[(1 - \pi_i^{1|0} - \pi_i^{0|1})(1 - \Phi(\beta\,'\mathbf{x}_i) + \pi_i^{0|1}\right]^{1-\tilde{y}_i}\Bigg],\end{aligned} \tag{5}$$

which represents the probability of observing the sample under misreporting. The joint posterior density of $\theta = \{\pi_i^{1|0}, \pi_i^{0|1}, \beta\,'\}$ is therefore given by:[13]

$$p(\beta, \pi_i^{1|0}, \pi_i^{0|1}|\mathcal{S}) \propto \mathcal{L}(\beta, \pi_i^{1|0}, \pi_i^{0|1}|\mathcal{S}) \times p(\beta, \pi_i^{1|0}, \pi_i^{0|1}). \tag{6}$$

Without prior substantive information, a common choice for $p(\pi_i^{1|0})$ and $p(\pi_i^{0|1})$ would be vague *Beta* distributions, while independent normal priors with zero means and (possible common) large variances could be assigned for the components of β (McInturff et al. 2004; Prescott and Garthwaite 2005). However, as mentioned above, using flat priors for the misclassification errors will likely lead to poor identifiability (Gu 2006). In addition, specifying diffuse priors for β can also hinder convergence in some circumstances (Gu 2006; McInturff et al. 2004; Prescott and Garthwaite 2002). Incorporating prior information on $\pi_i^{1|0}$, $\pi_i^{0|1}$ and β from auxiliary data sources can help overcome these problems and improve the accuracy of the parameter estimates (Gu 2006; McInturff et al. 2004; Prescott and Garthwaite 2002, 2005).

Suppose that both the true and the self-reported dependent variables are recorded for all respondents in a validation study of size M. Comparing y_j to $\tilde{y}_j$ for every $j = 1, \ldots, M$, we can estimate the misreport probabilities for the validated sample. Let $\mathbf{z}_j^1$ and $\mathbf{z}_j^2$ denote sets of regressors that

[12] This assumption simplifies the analysis considerably without entailing any obvious drawback from a practical perspective (Paulino, Soares and Neuhaus 2003).

[13] Alternatively, a "latent variable" approach based on data augmentation (Tanner and Wong 1987) can be used to simplify the computations. See McInturff et al. (2004).

are useful in predicting the conditional probabilities $\pi_j^{1|0}$ and $\pi_j^{0|1}$, where the notation allows for the fact we may use different regressors to predict the two types of misreporting. $\mathbf{z}_j^1$ and $\mathbf{z}_j^2$ may include some or all of the variables in $\mathbf{x}$, as well as other variables not affecting the true response. Again, for ease of exposition, we assume probit link functions and specify the conditional probabilities of misreporting as $\pi_j^{1|0} = \Phi(\gamma_1{'}\mathbf{z}_j^1)$ and $\pi_j^{0|1} = \Phi(\gamma_2{'}\mathbf{z}_j^2)$. Since our interest lies primarily on the distribution of β, $\gamma = \{\gamma_1^{'}, \gamma_2^{'}\}$ could in principle be viewed as "nuisance" parameters in our setting (Ibrahim, Ryan and Chen 1998), although they help provide meaningful interpretations for the underlying misreporting process (Chen 1979).[14] Letting $\mathcal{V}$ denote the data from the validation study, the likelihood from $\mathcal{V}$ is:

$$\begin{aligned}\mathcal{L}(\beta,\gamma_1,\gamma_2|\mathcal{V}) = &\prod_{j=1}^{M}\big(\Phi(\mathbf{x}_j)\big)^{y_j}\big(1-\Phi(\mathbf{x}_j)\big)^{(1-y_j)} \quad \times \\ &\prod_{y_j=1}\Phi(\gamma_1{'}\mathbf{z}_j^1)^{\tilde{y}_j}\times\big(1-\Phi(\gamma_1{'}\mathbf{z}_j^1)\big)^{1-\tilde{y}_j} \quad \times \\ &\prod_{y_j=0}\Phi(\gamma_2{'}\mathbf{z}_j^2)^{1-\tilde{y}_j}\times\big(1-\Phi(\gamma_2{'}\mathbf{z}_j^2)\big)^{\tilde{y}_j}.\end{aligned} \tag{7}$$

The posterior distributions $p(\beta,\gamma_1,\gamma_2|\mathcal{V})$ or $p(\gamma_1,\gamma_2|\mathcal{V})$ could then be used to specify the priors for β,γ_1 and γ_2 in the model fit to the sample of interest by repeated application of Bayes' theorem. However, since these posteriors cannot be expressed as tractable distributions, there is no straightforward way of transferring the relevant information from the validation study to the analysis of the main sample (Prescott and Garthwaite 2005). In addition, unless the validation study is a random sub-sample of the main study, heterogeneity between the two samples might in some circumstances lead to misleading conclusions if inference on β is based on the pooled datasets (Duan 2005). Hence, we consider both samples simultaneously, combining the likelihoods in 5 and 7 with vague independent priors $p(\beta), p(\gamma_1)$ and $p(\gamma_2)$ and weighting the likelihood from the validated sample by a "tunning" parameter δ that controls how much influence the validated data has relative to the main sample (Chen, Ibrahim and Shao 2000; Ibrahim and Chen 2000). The joint posterior density of the unknown parameters is therefore given by:

$$p(\beta,\pi_i^{1|0},\pi_i^{0|1}|\mathcal{S}) \propto \mathcal{L}(\beta,\pi_i^{1|0},\pi_i^{0|1}|\mathcal{S})\times\mathcal{L}(\beta,\gamma_1,\gamma_2|\mathcal{V})^{\delta}\times p(\beta)\times p(\gamma_1)\times p(\gamma_2) \tag{8}$$

with $0 \le \delta \le 1$, where $\delta = 0$ corresponds to the case in which no auxiliary information is incorporated into the analysis for the main sample, while $\delta = 1$ gives equal weights to $\mathcal{L}(\beta,\pi_i^{1|0},\pi_i^{0|1}|\mathcal{S})$ and $\mathcal{L}(\beta,\gamma_1,\gamma_2|\mathcal{V})$. δ can be assigned either a fixed value or a prior distribution – e.g., $a{\sim}Beta(c,d)$ – (Chen, Ibrahim and Shao 2000; Ibrahim and Chen 2000).[15] Although Equation 8 is intractable analytically, inference can be performed using Gibbs sampling along with Metropolis steps to sample the full conditionals for β, γ_1 and γ_2 (Gelfland and Smith 1990; Casella and George 1992; Chib

[14]It is worth mentioning, however, that $\pi^{1|0}(\mathbf{z}^1)$ and $\pi^{0|1}(\mathbf{z}^2)$ are not necessarily identified. See Lewbel (2000).

[15]In the latter case, the prior $p(\delta)$ would be added to Equation 8. See the discussions in Chen, Ibrahim and Shao (2000) and Ibrahim and Chen (2000) for additional details.

and Greenberg 1995). Under mild regularity conditions (Gilks, Richardson and Spiegelhalter 1996; Robert and Casella 2004), for a sufficiently large number of iterations, samples from these conditional distributions approach samples from the joint posterior. The posterior marginals obtained from these convergent samples can then summarized and used to estimate the effect of the relevant individual characteristics on the true response and the misreport probabilities. In addition, Bayes factors can be easily implemented within our modeling framework to compare alternative link functions (Paulino, Soares and Neuhaus 2003).

Thus, we only need to have validated data from a previous sample or for a sub-sample of the respondents in order to correct for misreporting in the model for the main study. In case several validation studies are available, they can be easily integrated into our analysis by adapting the method proposed in Ibrahim and Chen (2000) to incorporate historical data in binary choice models, substituting $\mathcal{L}(\beta, \gamma_1, \gamma_2|\mathcal{V})$ in Equation 8 by:

$$\prod_{d=1}^{D} \mathcal{L}(\beta, \gamma_1, \gamma_2|\mathcal{V}_d)^{\delta_d} \tag{9}$$

where $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_D\}$ denotes the data from D validation samples and $\delta = \{\delta_1, \ldots, \delta_D\}$, $0 \leq \delta_d \leq 1$ can be assigned I.I.D. beta priors (Ibrahim and Chen 2000; Ibrahim, Ryan and Chen 1998). Note that, while we must assume that the same error structure appears in the validated and non-validated samples and that the process generating misreporting is similar in both datasets, the covariates included in $\mathbf{x}$ and $\mathbf{z} = \{z^1, z^2\}$ do not have to be necessarily identical for both datasets. For instance, when estimating the determinants of the turnout decision, we could allow for election-specific factors affecting the turnout and the misreport probabilities, combining information from validation studies with experts' opinions, theoretical restrictions or even specifying diffuse priors for some of the predictors. Covariates that were not measured in previous studies can be incorporated into the analysis of the sample of interest by specifying the priors for these new covariates through the "initial" prior $p(\beta, \gamma_1, \gamma_2)$ in Equation 8 (Ibrahim et al. 2005).

Even if we did not have access to a validation sample, several other sources of information, such as administrative records or even aggregate data could be used to impose informative constraints on the misclassification rates and improve the parameter estimates. For example, in the analysis of voter turnout, we could observe turnout rates in small geographic areas, such as counties or congressional districts, that could be used to specify the misreport probabilities for all individuals in the sample belonging to a given area. While it will not be generally possible to specify a generalized linear model of misreporting in such circumstances, hierarchical beta priors can be used to summarize auxiliary information available on misreporting patterns by location or relevant socio-demographic characteristics (Dunson and Tindall 2000). Finally, if no relevant information to predict misreporting exists either in validation studies or other auxiliary data, constraints on the misreport probabilities could be imposed *via* elicitation of experts' opinions. Our model would then be virtually identical to McInturff et al. (2004) and Paulino, Soares and Neuhaus (2003).

Despite the advantages of our approach, it is worth mentioning that, like all parametric estimators, our model might be quite sensitive to distributional and modeling assumptions. Although semi-parametric methods have been used to estimate discrete choice models with misclassified dependent variables (Abrevaya and Hausman 1999; Hausman, Abrevaya and Scott-Morton 1998; Dustmant and van Soest 2004), they are also subject to potential misspecification (Molinari 2003). Moreover, in the case of covariate-dependent misclassification, available semi-parametric techniques require either sacrificing identification of some of the parameters in β (Abrevaya and Hausman 1999) or complex computations that are not likely to be attractive for practitioners and empirical researchers (Lewbel 2000). A different approach would be to adapt and implement non-parametric methods based on Manski (1985), Horowitz and Manski (1995) and Molinari (2003). In particular, the "direct misclassification approach" proposed by the latter allows incorporating prior information on the misreporting pattern to obtain interval identification of parameters of interest, and can be easily applied to the case in which misclassification depends on perfectly observed covariates with relatively little computational cost. However, as is well known, non-parametric methods are subject to the curse of dimensionality, which can pose a problem in applications where the misreporting probabilities might depend on a relatively large set of covariates, and is uncertain whether point identification can be achieved in this setting (Hu 2008). To the best of our knowledge, there is very little research comparing the performance of parametric *versus* non-parametric methods to correct for covariate-dependent misclassification and evaluating the relative weaknesses and advantages of both approaches in applied work.

1.3. *Extending the model to account for missing data*

Besides measurement errors, survey data is often plagued with large proportions of missing outcome and covariate values due to non-response or loss of data. As is well known, unless the data are missing completely at random (MCAR), using list-wise deletion and restricting the analysis only to those respondents who are completely observed can lead to biased estimates (Little and Rubin 2002; Chen et al. 2008).[16] Furthermore, even if the data are MCAR, complete-case analyses may lead to discard a large proportion of observations and can be therefore quite inefficient (Ibrahim et al. 2005). Ad-hoc approaches to dealing with missing data, such as excluding covariates subject to missingness from the analysis or using mean imputation, are easy to implement but exhibit several potential problems such as biased estimates, inefficiency and misspecification (Chen et al. 2008; Ibrahim et al. 2005; Gelman and Hill 2007).[17] On the other hand, Bayesian methods such as the one presented in this paper can easily accommodate missing data. There is no distinction between missing data and parameters within the Bayesian framework, and thus inference in this setting essentially requires defining a prior for the missing values and sampling from the joint posterior distribution of the parameters and missing values, incorporating just an "extra-layer" in the Gibbs sampling algorithm compared to the complete-case analysis (Gelman et al. 2004; Ibrahim et al. 2005). In particular, our model can be immediately extended to deal with missing response and covariate values, including cases with missing responses alone, with missing covariates alone, and with missing covariates and responses. This allows us to accommodate item and unit nonresponse in both the main and the validation studies.[18]

Let $\mathbf{w}_i = (w_{i,1}, \ldots, w_{i,p})'$, $i = 1, \ldots, N$, denote a $p \times 1$ vector of covariates included in $\mathbf{x}_i, \mathbf{z}_i^1$ and $\mathbf{z}_i^2$, and denote the marginal density of $\mathbf{w}_i$ by $p(\mathbf{w}_i|\alpha)$, where α parametrizes the joint distribution of the covariates. Adopting the notation in Chen et al. (2008), we write $\mathbf{w}_i = (\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis})$,

[16] It is worth mentioning, however, that there are situations in which inference based on a complete-case analysis might yield unbiased estimates and outperform imputation methods even when the data are not missing completely at random (Little and Wang 1996).

[17] A detailed review of different methods commonly used to handle missing data is beyond the scope of this paper. See Horton and Kleinman (2007), Ibrahim et al. (2005), Little and Rubin (2002) and Schafer and Graham (2002), among others, for a detailed discussion.

[18] However, as seen in Equation 11 below, respondents with completely missing outcomes and covariates do not contribute to the likelihood function.

where $\mathbf{w}_{i,mis}$ is the $q_i \times 1$ vector of missing components of $\mathbf{w}_i$, $0 \leq q \leq p$, and $\mathbf{w}_{i,obs}$ is the observed portion of $\mathbf{w}_i$. Similarly, we use $\tilde{y}_{i,mis}$ if the self-reported outcome $\tilde{y}_i$ is missing, and $\tilde{y}_{i,obs}$ otherwise. We assume that the missing data mechanism is *ignorable* (Rubin 1976; Little and Rubin 2002). That is, we assume that the missing data mechanism does not depend on the missing values, but may depend on the observed outcome and covariate data included in the model – i.e., the data are missing at random (MAR) – and that the parameters governing the missing data mechanism are distinct from the parameters of the sampling model. The *observed-data likelihood* for the main study can then be written as:

$$
\begin{aligned}
\mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha | \mathcal{S}_{obs}) = & \prod_{\tilde{y}_{i,obs}, \mathbf{w}_i = \mathbf{w}_{i,obs}} p(\tilde{y}_i | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_i | \alpha) \quad \times \\
& \prod_{\tilde{y}_{i,obs}, \mathbf{w}_i = (\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis})} \int p(\tilde{y}_i | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis} | \alpha)\, d\mathbf{w}_{i,mis} \quad \times \\
& \prod_{\tilde{y}_{i,mis}, \mathbf{w}_i = \mathbf{w}_{i,obs}} \int p(\tilde{y}_{i,mis} | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_i | \alpha)\, d\tilde{y}_{i,mis} \times \\
& \prod_{\tilde{y}_{i,mis}, \mathbf{w}_i = (\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis})} \int \int p(\tilde{y}_{i,mis} | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis} | \alpha)\, d\tilde{y}_{i,mis}\, d\mathbf{w}_{i,mis} \times \\
& \prod_{\tilde{y}_{i,mis}, \mathbf{w}_i = \mathbf{w}_{i,mis}} \int \int p(\tilde{y}_{i,mis} | \mathbf{w}_{i,mis}, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_{i,mis} | \alpha)\, d\tilde{y}_{i,mis}\, d\mathbf{w}_{i,mis},
\end{aligned}
\tag{10}
$$

which, as noted by Chen et al. (2008), reduces to:

$$
\begin{aligned}
\mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha | \mathcal{S}_{obs}) = & \prod_{\tilde{y}_{i,obs}, \mathbf{w}_i = \mathbf{w}_{i,obs}} p(\tilde{y}_i | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_i | \alpha) \quad \times \\
& \prod_{\tilde{y}_{i,obs}, \mathbf{w}_i = (\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis})} \int p(\tilde{y}_i | \mathbf{w}_i, \beta, \gamma_1, \gamma_2)\, p(\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis} | \alpha)\, d\mathbf{w}_{i,mis} \quad \times \\
& \prod_{\tilde{y}_{i,mis}, \mathbf{w}_i = \mathbf{w}_{i,obs}} p(\mathbf{w}_i | \alpha) \ \times \\
& \prod_{\tilde{y}_{i,mis}, \mathbf{w}_i = (\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis})} \int p(\mathbf{w}_{i,obs}, \mathbf{w}_{i,mis} | \alpha)\, d\mathbf{w}_{i,mis}.
\end{aligned}
\tag{11}
$$

As suggested by Ibrahim, Chen and Lipsitz (2002), it is often convenient to model the joint distribution $p(\mathbf{w}_i|\alpha)$ is as a series of one-dimensional conditional distributions:

$$
\begin{aligned}
p(w_{i,1}, \ldots, w_{i,p} | \alpha) = & \, p(w_{i,p} | w_{i,1}, \ldots, w_{i,P-1}, \alpha_p) \\
& \times p(w_{i,p-1} | w_{i,1}, \ldots, w_{i,p-2}, \alpha_{p-1}) \times \cdots \times p(w_{i,1} | \alpha_1)
\end{aligned}
\tag{12}
$$

where α_l, $l = 1, \ldots, p$, is a vector of parameters for the lth conditional distribution, the α_l's are distinct, and $\alpha = (\alpha_1, \ldots, \alpha_p)$. As noted by these authors, specification 12 has the advantages of easing the prior elicitation for α and reducing the computational burden of the Gibbs algorithm required for sampling from the observed data posterior, and is particularly well-suited for cases in which $\mathbf{w}$ includes categorical and continuous covariates. While the modeling of the covariate distributions depends on the order of the conditioning, Ibrahim, Chen and Lipsitz (2002) show that posterior inferences are generally quite robust to changes in the order of the conditioning. Obviously, 12 needs to be specified only for those covariates that have missing values. If some of the covariates in w are completely observed for all respondents in a survey, they can be conditioned on when constructing the distribution of the missing covariates.

The joint posterior density of the unknown parameters based on the observed data is then given by:

$$p(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{S}_{obs}) \propto \mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{S}_{obs}) \times p(\beta, \gamma_1, \gamma_2, \alpha) \tag{13}$$

Information on the misreport patterns and on all the parameters of interest can be incorporated from the validation study in essentially identical way as in the case with no missing data. A joint prior for $(\beta, \gamma_1, \gamma_2, \alpha)$ could be specified as:

$$p(\beta, \gamma_1, \gamma_2, \alpha) \propto \mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{V}_{obs})^{\delta} \times p(\beta) \times p(\gamma_1) \times p(\gamma_2) \times p(\alpha), \tag{14}$$

where $\mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{V}_{obs})$ is obtained from the complete-data likelihood of the validation study:

$$\mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{V}_{obs}) = \int\int p(\tilde{\mathbf{y}}, \mathbf{y}|\mathbf{w}, \beta, \gamma_1, \gamma_2, \alpha)\, d\tilde{y}_{mis}\, d\mathbf{w}_{mis}, \tag{15}$$

and, as mentioned in 1.2, δ is a scalar prior parameter that weights the validated data relative to the data from the main study.[19] Note that our specification allows for missing responses $\tilde{y}_i$ and covariate values in the validated sample as well, and can accommodate cases in which the missing self-reported variable depends on the true y_i. As in the case of no missing data, it is also possible to incorporate only the information from the observed probability of misreporting in the validation study to specify the priors for γ_1, γ_2 and a subset α_z of the components of α for the main study, while using diffuse prior distributions for the remaining parameters. However, the additional information obtained from $\mathcal{L}(\beta, \gamma_1, \gamma_2, \alpha|\mathcal{V}_{obs})$ can increase efficiency in many missing data problems in which certain parameters in the likelihood function are not identifiable and/or very little information is available for inference, particularly when the "gold-standard" measure y_i

[19]See Section 4 in Ibrahim, Chen and Lipsitz (2002) for details.

is observed for a large proportion of the respondents in the validation study (Ibrahim, Chen and Lipsitz 2002; Reilly and Pepe 1995; Robins, Rotnitzky and Zhao 1994).

In principle, it is possible to extend this approach to the case of non-ignorably missing values following Huang, Chen and Ibrahim (1999), Ibrahim and Lipsitz (1996) and Ibrahim, Lipsitz and Chen (1999). However, there is usually little information on the missing data mechanism, and the parameters of the missing data model are often quite difficult to estimate (Ibrahim, Lipsitz and Horton 2001). The plausibility of the MAR assumption can be enhanced by including additional individual and contextual variables in the model specification (Gelman et al. 2004; Gelman, King and Liu 1998).

2. A MONTE CARLO EXPERIMENT

In this section, we conduct a series of simulation analyses aimed at illustrating the problems of ignoring misreporting in practice, comparing the performance of our solution *vis a vis* alternative parametric models proposed in the literature to account for misreporting, and assessing the sensitivity of the estimates from our model to the specification of the underlying model of misreporting.

2.1. *Comparison of alternative approaches to dealing with misreporting*

Based on the Monte Carlo design in Neuhaus (1999), we simulated 2,000 observations for two covariates: x_1 is drawn from a standard normal distribution, and x_2 is a dummy variable equal to one with probability 1/2. The true response y_i was generated as:

$$y_i = I(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i \geq 0)$$

where $I(E)$ is the indicator function equal to one if E is true and zero otherwise, $(\beta_0, \beta_1, \beta_2) = (-1, 1, 1)$ and ϵ_i drawn from a $N(0, 1)$ distribution.

The misreport probabilities $\pi_i^{1|0}$ and $\pi_i^{0|1}$ were chosen such that: i) average misclassification rates are symmetric and take values of of 2%, 5%, 10% and 20%; ii) different possible relationships between $\mathbf{x}_i$, $\pi_i^{1|0}$ and $\pi_i^{0|1}$ are taken into account. This allows us to determine whether and to what extent ignoring misclassification affects the parameter estimates for different rates of misreporting and for different correlation patterns between the covariates of interest, the true response and the misreport probabilities. For reasons of space, we only present the results for the two basic scenarios considered by Neuhaus (1999), denoted as Designs A and B.[20] In Desing A, $\pi_i^{1|0}$ and $\pi_i^{0|1}$ are independent of the covariates in $\mathbf{x}$; the observed response $\tilde{y}_i$ was generated by randomly changing y_i according to the constant misreport probabilities. Under Design B, the binary covariate x_2 is assumed to be stronlgy positively correlated with $\pi^{1|0}$ but negatively related to $\pi^{0|1}$; $\tilde{y}_i$ in this scenario was generated from y_i as a function of x_2, as indicated in Table 3 in Appendix A. This corresponds, for instance, to the situation described in previous analysis of voter turnout that found overreporting to be clearly correlated with race (Abramson and Claggett 1984, 1986*a*,*b*, 1991; Hill and Hurley 1984; Sigelman 1982). For all simulated datasets, we impose the monotonicity condition $\pi_i^{1|0} + \pi_i^{0|1} < 1$ (Hausman, Abrevaya and Scott-Morton 1998).

In order to apply the methodology developed in Section 1.2, we randomly selected half of the observations in the sample and assigned them to be the validation study. The remaining

[20]Simulations were also carried out allowing for $\pi_i^{1|0} \neq \pi_i^{0|1}$. The results for the entire set of simulation exercises are available from the authors upon request.

1,000 observations were assigned to be the main sample under analysis, and we ignored the true response and the information on the misclassification probabilities for these observations, using the information on conditional misreport probabilities from the validation study to fit the model in Equation 8 with x_1 and x_2 as regressors both in the response and the misreport models.Since the validation study is a random sub-sample of the main study, a point mass prior $\delta = 1$ with probability 1 was used, equally weighting the validated and main samples.[21] It is worth noting that, following Equation 7, we adopt a probit specification for the misreport probabilities, despite the fact that using a probit link to estimate a binary choice model with nonnormal error terms can yield biased and inconsistent parameter estimates (Horowitz 1993). However, as mentioned above, the main purpose of our method is to improve inferences on the conditional distribution of the true response given some covariates of interest rather than to estimate the conditional misreport probabilities. Adopting a probit specification commonly used by practitioners for the underlying misreport model allows us to assess how robust are the estimates of β to common misspecification errors likely to emerge in applied work.[22] In 2.3 we examine the sensitivity of our method to various forms of misspecification of the misreport model in more detail.

We compared the estimates from our method with those obtained using a standard probit model ignoring misreporting, as well as from two alternative approaches proposed in the literature to correct for misclassification. Model A-1 is similar to Hausman, Abrevaya and Scott-Morton (1998)'s parametric estimator, assuming constant misclassification probabilities and ignoring the information from the auxiliary data. Model A-2 also assumes covariate-independent misreporting, but information on the posterior distribution of $\pi^{1|0}$ and $\pi^{0|1}$ from the validation study is used to define $p(\pi_i^{1|0})$ and $p(\pi_i^{0|1})$ for the main sample, as suggested in Prescott and Garthwaite (2002, 2005). All models were fit *via* MCMC methods, assigning independent $N(0, 100)$ priors for the components of β. The parameters in γ_1 and γ_2 under model were assigned independent $N(0, 100)$ distributions, while independent $Beta(c, d)$ priors were specified for $\pi_i^{1|0}$ and $\pi_i^{0|1}$ under Models A-1 and A-2. c and d were set equal to 1 for Model A-1 and extracted from the misclassification rates in the validated sample for Model A-2.[23] The models were fit in WinBUGS 1.4. Three parallel chains of length 50,000 with over-dispersed initial values and a 5,000 period burn-in were run for each model; convergence was assessed based on Gelman and Rubin's estimated Potential Scale Reduction Factor (Gelman and Rubin 1992).

Tables 4 and 5 in Appendix A report the posterior means and central 95% credible intervals for all the parameter estimates from the four different estimation approaches, and Figure 1 below plots the approximate posterior density for the coefficient of the binary covariate x_2. As the amount of misclassification increases, the point estimates (means and medians) for all the parameters under the standard probit specification become further away from the true values. The central 95% credible intervals from the model ignoring misclassification fail to cover the true $(\beta_0, \beta_1, \beta_2)$ for average misreport probabilities larger than 5% under Design A, as well as the true coefficients of the simulated covariates under Design B. However, even for average misclassification rates $\bar{\pi}^{1|0}$ and $\bar{\pi}^{0|1}$ as low as 0.05, the true coefficient of x_1 lies outside the credible intervals from the probit model under both experimental designs. In contrast, under our proposed estimation solution, the point estimates for all the parameters are much closer to the true β and the central 95% intervals cover

[21] Changes in the values of δ have relatively little effect on the parameter estimates in our setting. See the results in Section 3.3.

[22] Clearly, researchers should also worry about misspecified response models, but this problem is common to all parametric binary choice models.

[23] See Section 2 in Prescott and Garthwaite (2002).

them for all values of $\bar{\pi}^{1|0}$ and $\bar{\pi}^{0|1}$ under the two simulation scenarios. Similar results (not shown) are obtained for the ratios of the estimated coefficients of the simulated covariates with respect to the intercept. We also note that the standard deviations for the estimates under our model tend to be larger than for the simple probit model due to the fact that our model captures the additional uncertainty in the true latent variable y_i induced by misreporting (McGlothlin, Stamey and Seaman 2008; Neuhaus 1999). For the same reason, the standard deviations from our model increase considerably with the amount of misclassification.

A comparison with the two alternative approaches to correct for misreporting shows that, when $\pi_i^{1|0}$ and $\pi_i^{0|1}$ are assumed constant, the estimates from our method do not differ substantially from those obtained under Model A-2. The widths of the central 95% credible intervals are also similar for both models, even though we might have expected that the stronger distributional assumptions about the misreport process adopted in our approach should lead to narrower intervals (Prescott and Garthwaite 2005). In the case of covariate-dependent misreporting, however, the performance of Model A-2 worsens markedly. In particular, as seen Figure 1, the posterior mean for β_2 becomes implausibly large as the misclassification rates and the correlation between x_2 and the misreport probabilities increase, and the point estimates for the intercept also become far away from the true β_0. Model A-1, on the other hand, fails to converge for all $\bar{\pi}^{1|0}$, $\bar{\pi}^{0|1} > 0.02$ under Design B, and performs much worse than our model and Model A-2 also in the case of constant misclassification rates. Furthermore, as seen in Table 6 in Appendix A, Model A-1 yields markedly biased and imprecise estimates of the average misclassification rates in the main sample, indicating that assigning diffuse distributions for $p(\pi_i^{1|0})$ and $p(\pi_i^{0|1})$ when the data provides very little information to estimate the misreport probabilities results in very volatile and inaccurate estimates for the parameters of interest, even when the Hausman, Abrevaya and Scott-Morton (1998)' identification condition $\pi_i^{1|0} + \pi_i^{0|1} < 1$ holds for all the cases. In contrast, the point and interval summaries of the posterior distributions under our proposed modle and under Model A-2 are quite similar in most cases, although the credible intervals from the latter fail to cover the true average misreport rates in the scenario with covariate-dependent misreporting for $\bar{\pi}^{1|0}$, $\bar{\pi}^{0|1} = 20\%$.

In order to illustrate the differences between inferences based on the alternative estimators, Figure 2 plots the estimated marginal effect of x_1 and x_2 on the probability that the response takes a value of 1 under the standard probit model ignoring misreporting, our proposed method, and the approach based on Prescott and Garthwaite (2002, 2005) (Model A-2). For both simulated covariates, the true average effects estimated using y_i as the dependent variable always lie comfortably within the central 95% credible intervals from our model under both Monte Carlo designs and for all the misclassification rates considered. The maximal differences between the point estimates from our model and the true effects are at most of 2 and 8 percentage points for x_1 and x_2, respectively. In contrast, the "naive" probit model systematically underestimates the marginal effect of x_1 in the two simulation scenarios and leads to strongly biased estimates for the effect of x_2 for $\bar{\pi}^{1|0}$, $\bar{\pi}^{0|1} \geq 10\%$. The differences in the performance of the two models increase with the prevalence of misreporting and are most notorious in the case in which $\pi_i^{1|0}$ and $\pi_i^{0|1}$ depend on x_2. For $\bar{\pi}^{1|0}$,$\bar{\pi}^{0|1} = 0.2$, the effect of x_1 estimated without adjusting for misclassification is less than half the true value, and the marginal effect of x_2 is overestimated by more than 30 percentage points. Also, while the marginal effects estimated using our approach and Model A-2 are quite similar under the scenario with constant misclassification rates, our model performs much better than Model A-2 when misreporting is covariate-dependent. In fact, for the binary covariate x_2, ignoring misreporting yields more accurate estimates of the marginal effects than incorporating information from the validation sample in the way suggested by Prescott and Garthwaite (2002,

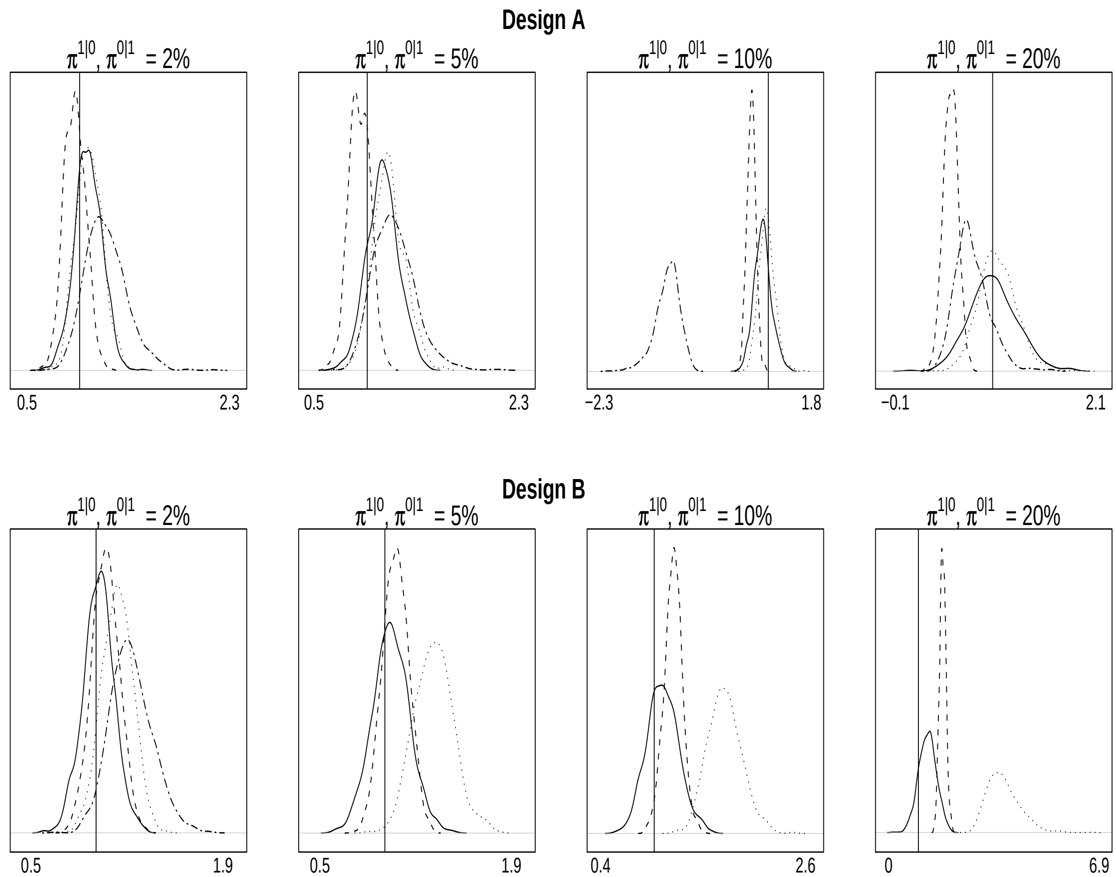

Figure 1: **Estimated posterior densities for β_2 across models.** The graph compares marginal posterior density for β_2 under four different estimation approaches: our proposed method (solid curve), a probit model ignoring misreporting (dashed curve), Model A-1 (double-dashed curved) and Model A-2 (dotted curve). The solid vertical line denotes the true parameter value.

2005).

We also fit our model using several different specifications for the model of misreporting and found again that the point estimates of $\beta = \{\beta_0, \beta_1, \beta_2\}$ and the marginal covariate effects under our approach are closer to the true values than under any of the alternative methods considered. Figure 3 illustrates this, plotting the marginal effects of x_1 and x_2 for Design B and $\bar{\pi}^{1|0}$, $\bar{\pi}^{0|1} = 0.2$ under three alternative specifications for the linear predictor of the misreport model: adding an interaction term between x_1 and x_2 , omitting x_1, and including an additional variable that is not significantly related to either the misreport probabilities or the true response. The point estimates differ slightly across specifications and are more accurate for the average effect of the normally distributed covariate. Nonetheless, a comparison with the results in Figure 2 shows that our method performs considerably better than the other estimators under all the specifications of the misreport model considered.

Hence, the evidence from this simulation study shows that, in the presence of misreporting, our method can considerably improve the accuracy of the parameter estimates with respect to standard binary choice models even for misclassification rates as low as 5%. When misreporting is

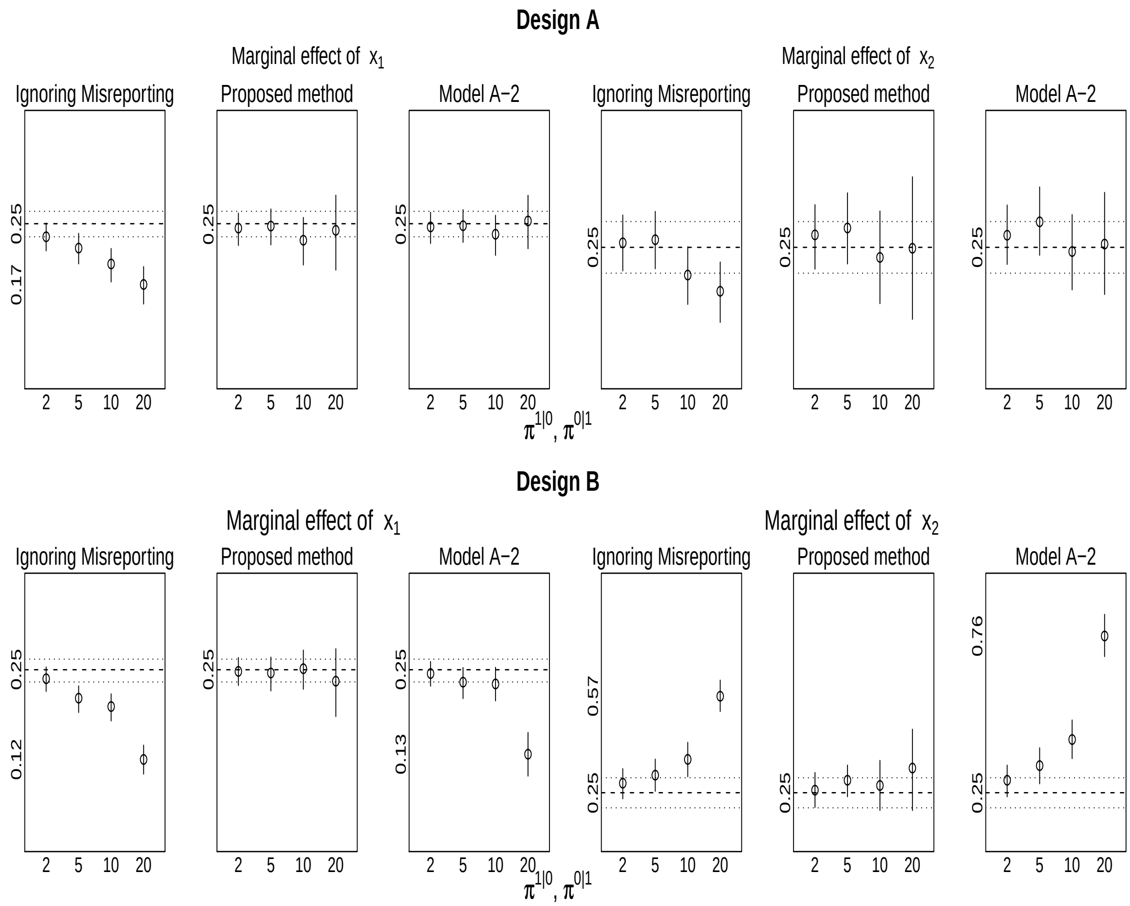

Figure 2: **Marginal covariate effects.** The graph compares the marginal effects of the two simulated covariates x_1 and x_2 estimated under three different approaches: a standard probit model, our method correcting for misreporting, and Model A-2. The center dots correspond to the the posterior means, the vertical lines to the central 95% credible intervals, and the horizontal lines represent the average effects (dashed) and 95% intervals (dotted) estimated using y_i as the response.

covariate-dependendent, our proposed estimation solution also performs considerably better than alternative approaches assuming constant misclassification rates, especially for non-trivial levels of misclassification. Differences in the parameter estimates obtained under alternative models may consideraly affect inferences drawn from the sample under analysis, and thus ignoring misreporting or neglecting the potential correlation between the covariates of interest and the misreport probabilities could lead to quite different substantive conclusions. In addition, our results indicate that using auxiliary information may be critical to improve identifiability and convergence properties of models correcting for misclassification in relatively small samples, such as those typically used in political science.

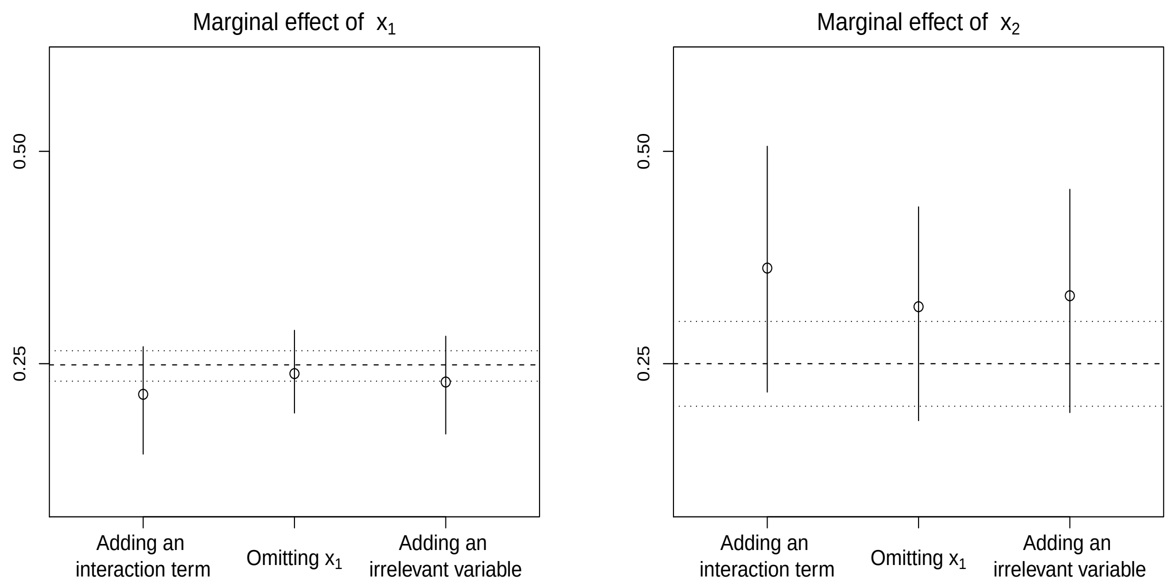

Figure 3: ***Marginal covariate effects under different specifications of the misreport model.*** *The graph compares the marginal effects of x_1 and x_2 estimated under our proposed method, using three alternative specifications of the linear predictor in the model of misreporting. The center dots correspond to the the posterior means, the vertical lines to the central 95% credible intervals, and the horizontal lines represent the average effects (dashed) and 95% intervals (dotted) estimated using y_i as the response.*

2.2. *Assessing robustness to the specification of the misreport model*

The results from the simulation study reported in 2.1 indicate that the model proposed in this paper can successfully adjust for misreporting under different parametric models for the misclassification mechanism. Nonetheless, the sensitivity of our method to misspecification of the model of misreporting deserves further attention, since this may lead to inconsistent estimates β and affect inferences on the covariate of interest (Abrevaya and Hausman 1999; Hausman, Abrevaya and Scott-Morton 1998). In order to examine this issue in more detail, we draw on research analyzing a somewhat similar problem, namely, the sensitivity of the estimated treatment effects to the specifications of the propensity score model (Drake 1993; Zhao 2008).

Our main goal here is to examine the influence on the estimated covariate effects of misspecifying the disturbance distribution in the model of misreporting, omitting relevant covariates from the linear predictor, including variables not related to either the true response or the misreport probabilities and adding unnecessary nonlinear terms. Specifically, using the covariates and the true response from 2.1, we generate a dichotomous variable d_i as:

$$d_i = \begin{cases} I\big(\gamma_{1,0} + \gamma_{1,1}x_{i,1} + \gamma_{1,2}x_{i,2} + \eta_i \geq 0\big); & \text{if } y_i = 0 \\ I\big(\gamma_{2,0} + \gamma_{2,1}x_{i,1} + \gamma_{2,2}x_{i,2} + \eta_i \geq 0\big); & \text{if } y_i = 1 \end{cases}$$

η is an error term, and $\gamma_1 = \{\gamma_{1,0}, \gamma_{1,2}, \gamma_{1,3}\}$, $\gamma_2 = \{\gamma_{2,0}, \gamma_{2,2}, \gamma_{2,3}\}$, are chosen to obtain different levels of misclassification and different degrees of correlation between the simulated covariates and the misreport probabilities $\pi_i^{1|0}$ and $\pi_i^{0|1}$. The observed response $\tilde{y}_i$ is in then generated as:

$$\tilde{y}_{ii} = \begin{cases} I(d_i = 1); & \text{if } y_i = 0 \\ 1 - I(d_i = 1); & \text{if } y_i = 1. \end{cases}$$

In order to analyze the sensitivity of our method to misspecification of the error disturbance in the model of misreporting, we follow Horowitz (1993); Drake (1993); Zhao (2008) and consider 4 distributions for η: a standard normal distribution, a logistic distribution, a bimodal distribution $\eta = 0.5N(3,1) + 0.5N(-3,1)$, and heteroskedastic error terms $\eta \sim N(1, 1 + 0.1x_1^2)$. We also implement 4 alternative specifications for the linear predictor of the misreport model:

$$\begin{aligned} &\text{Specification 1}: && \alpha_{k,0} + \alpha_{k,1} x_{i,2}; \\ &\text{Specification 2}: && \alpha_{k,1} x_{i,1} + \alpha_{k,2} x_{i,2} + \alpha_{k,3} x_{i,1}^2; \\ &\text{Specification 3}: && \alpha_{k,0} + \alpha_1 x_{i,1} + \alpha_{k,2} x_{i,2} + \alpha_{k,3}(x_{i,1} \times x_{i,2}); \\ &\text{Specification 4}: && \alpha_{k,0} + \alpha_{k,1} x_{i,1} + \alpha_{k,2} x_{i,2} + \alpha_{k,3} x_{i,3}; \end{aligned}$$

with $k = 1, 2$ and x_3 drawn from a log-normal distribution. We examine the effect of both forms of misspecification separately - i.e., we correctly specify the linear predictor of the misreport model when analyzing the role of misspecified error distributions and use standard normal errors when examining the influence of the functional form of the index term.[24]

Figure 4 reports the estimates of the marginal covariate effects when x_1 is omitted from the linear predictor of the misreport model (Specification 1) for different values of $\gamma_{1,1}, \gamma_{2,1}$ and average symmetric misreport rates of approximately 10% and 20%.[25] The estimates of the marginal effect of x_1 worsen as the average misclassification rates increase and as the the correlations between the covariate and the misreport probabilities increase. However, for all values of $\gamma_{1,1}, \gamma_{2,1}$, the estimates from our model are closer to the true marginal effects obtained using the true data than the estimates from a model ignoring misreporting. The estimates for x_2, on the other hand, are not affected by the omission of x_1 from the model of misreporting and are again much more accurate than those from a standard probit model.

Table 1 complements the information from the figure, illustrating the influence of the other forms of misspecification considered for different values of $\bar{\pi}^{1|0}, \bar{\pi}^{0|1}, \gamma_1$ and γ_2. In line with the results in 2.1, adding irrelevant covariates and unnecessary nonlinear terms to the linear predictor of the misreport model has relatively little influence on the estimated marginal effects, and the same holds for the case of misspecified disturbance distributions. In all cases, the true average covariate effects lie within the central 95% credible intervals from our model, and the point estimates are betwen 4 and 18 percentage points closer to the true values than the those obtained ignoring misreporting. Note that, as indicated in Tables 7 and 8 in Appendix A, the estimates of γ_1 and γ_2 can be far away

[24]We also let the covariates in $\mathbf{z}^1$ and $\mathbf{z}^2$ differ across specifications and consider several values of γ_1, γ_2, with little change in the main substantive results presented in this section.

[25]In all cases, we set $\gamma_{1,2} = 1.25$, $\gamma_{2,2} = -1.25$, and adjust the value of the intercept to achieve the desired average misclassification rates.

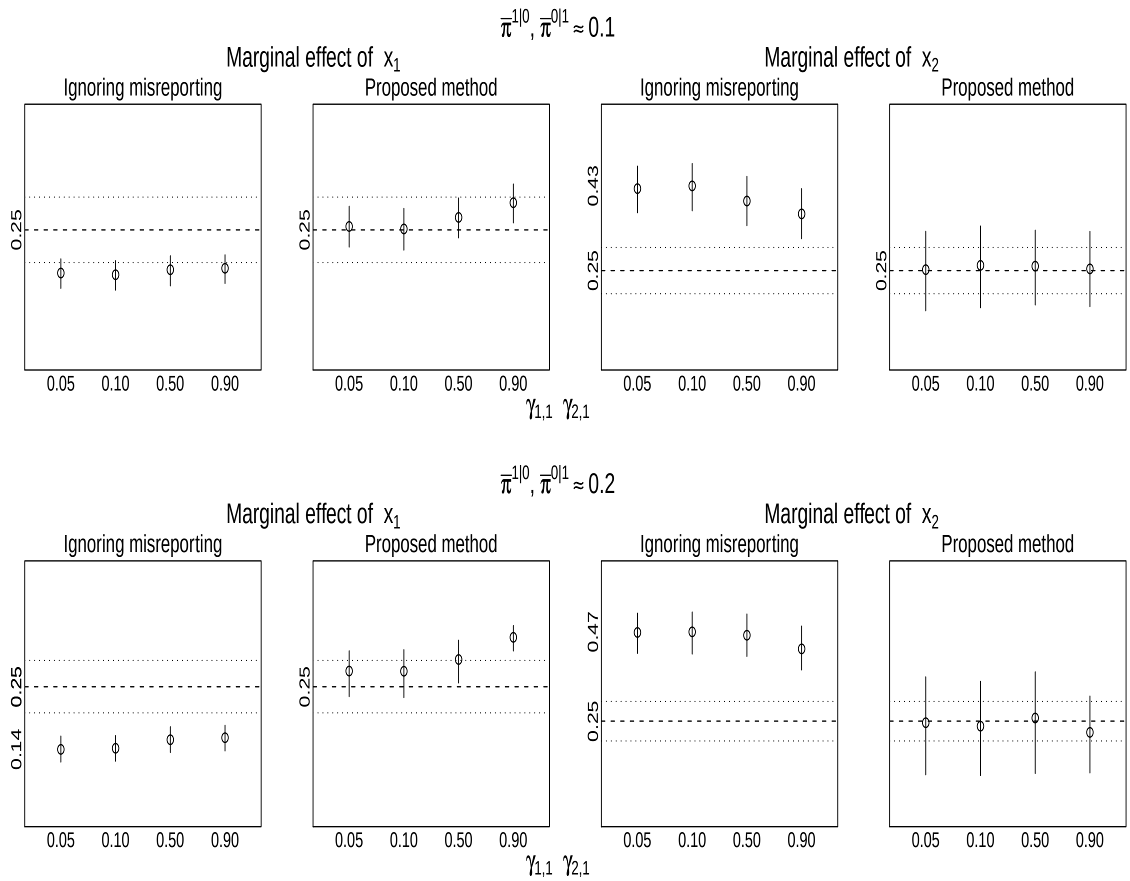

Figure 4: ***Marginal covariate effects when** x_1 **is omitted from the misreport model.*** *The graph plots the marginal effects* x_1 *and* x_2 *estimated under our method when* x_1 *is omitted from the linear predictor of the misreport model, for different values of* γ_1 *and* γ_2*. The center dots correspond to the the posterior means, the vertical lines to the central 95% credible intervals, and the horizontal lines represent the average effects (dashed) and 95% intervals (dotted) estimated using* y_i *as the response.*

from the true coefficients when the model of misreporting is misspecified, particularly when the error terms are bimodal or heteroskedastic (Horowitz 1993; Zhao 2008). However, the estimated covariate effects seem to be quite robust to the specification of the misreport model and much more accurate than those from standard parametric models when misclassificaton is non-negligible. We must note, though, that these results are based on limited simulation analyses and may not be true in general.

Table 1: Marginal covariate effects under alternative specifications of the misreport model

	Estimator	$\partial Pr(y=1\|\mathbf{x})/\partial x_1$	$\partial Pr(y=1\|\mathbf{x})/\partial x_2$
	True Model	0.25 (0.24, 0.27)	0.25 (0.21, 0.30)
Linear predictor[a]			
	Specification 2	0.29 (0.24, 0.33)	0.26 (0.13, 0.38)
	Specification 3	0.29 (0.24, 0.32)	0.27 (0.15, 0.38)
	Specification 4	0.29 (0.24, 0.33)	0.26 (0.15, 0.38)
Error disturbance			
	Logistic distribution[b]	0.27 (0.21, 0.32)	0.23 (0.09, 0.38)
	Bimodal distribution[c]	0.24 (0.20, 0.28)	0.21 (0.10, 0.29)
	Heteroskedastic[d]	0.24 (0.17, 0.29)	0.28 (0.17, 0.39)
Different misreport models in both sub-samples[e]		0.24 (0.21, 0.28)	0.30 (0.21, 0.38)

[a] $\gamma_{1,0} = -1.5, \gamma_{1,1} = 0.05, \gamma_{1,2} = 1.25, \gamma_{2,0} = -0.2, \gamma_{2,1} = 0.05, \gamma_{2,2} = -1.25, \bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.2.$

[b] $\gamma_{1,0} = -1.75, \gamma_{1,1} = 0.65, \gamma_{1,2} = 1.3, \gamma_{2,0} = -0.75, \gamma_{2,1} = 0.20, \gamma_{2,2} = -1.3, \bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.2.$

[c] $\gamma_{1,0} = -1.6, \gamma_{1,1} = 0.5, \gamma_{1,2} = 1.3, \gamma_{2,0} = -1, \gamma_{2,1} = 0.5, \gamma_{2,2} = -1.30, \bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.1.$

[d] $\gamma_{1,0} = -2.05, \gamma_{1,1} = 0.95, \gamma_{1,2} = 0.1, \gamma_{2,0} = -1.5, \gamma_{2,1} = -2.5, \gamma_{2,2} = -0.70, \bar{\pi}^{1|0} \approx 0.1, \bar{\pi}^{0|1} \approx 0.2.$

[e] $\bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.1$

Validation sample: $\gamma_{1,0} = -1.8, \gamma_{1,1} = 0.52, \gamma_{1,2} = 1.3, \gamma_{2,0} = -1.1, \gamma_{2,1} = 0.5, \gamma_{2,2} = -1.3.$

Main sample: $\gamma_{1,0} = -2.14, \gamma_{1,1} = 0.89, \gamma_{1,2} = 1.74, \gamma_{2,0} = -1.22, \gamma_{2,1} = 0.76, \gamma_{2,2} = -1.32.$

We also conducted additional simulations assuming a slightly different misreport processes for the validated and the main samples. Specifically, the values of γ_1 and γ_2 in the main sample were obtained by adding uniformly distributed errors to the corresponding parameters from the validation study, preserving the amount of misclassification and the direction of the relationship between the covariates and the misreport probabilities but changing the magnitude of the effect of x_1 and x_2 on $\pi_i^{1|0}$ and $\pi_i^{0|1}$. Again, as illustrated at the bottom of Table 1, the marginal effects estimated from our model are quite close to the true covariate effects. In contrast, the model ignoring misclassification systematically underestimates $\partial Pr(y = 1|\mathbf{x})/\partial x_1$ and overestimates $\partial Pr(y = 1|\mathbf{x})/\partial x_2$.

2.3. *Accounting for missing response and covariate values*

Finally, we compared the performance of our proposed method to other approaches in the presence of both misclassification and missing data. For this exercise, we draw x_1 from a standard normal distribution, as in 2.1, and simulate x_2 from a Bernoulli distribution with success probability modeled as $Pr[x_{i,2} = 1] = \Phi(\phi_0 + \phi_1 |x_{i,1}|)$. We assume that $x_{i,1}$ is completely observed for all subjects, and that $x_{i,2}$ and the observed response $\tilde{y}_i$ are missing at random (MAR) for some subjects. The missing mechanisms for $\tilde{y}_i$ and $x_{i,2}$ are:

$$\Pr[\mathrm{m}_i^{\tilde{y}} = 1] = \Phi(\alpha_{1,0} + \alpha_{1,1} x_{i,1} + \alpha_{1,2} x_{i,2}) \qquad \text{and}$$

$$\Pr[\mathrm{m}_i^{x_2} = 1] = \Phi(\alpha_{2,0} + \alpha_{2,1} x_{i,1});$$

where $m_i^{\tilde{y}} = 1$ or $m_i^{x_2} = 1$ if $\tilde{y}_i$ or $x_{i,2}$ is observed, and 0 otherwise. Using the same Monte Carlo designs as in 2.1, we generated samples of 2,000 observations with various levels of misclassification and different patterns of missing covariates and response, ignoring the true response y_i for half of the sample.

Table 2 illustrates the results for two combinations of misreporting and missing data patterns, contrasting the estimates of β from our method with those from a probit model ignoring misclassification and from Model A-2, based on Prescott and Garthwaite (2002, 2005).[26] For the three estimators, we use a fully Bayesian approach for inference with missing covariate and response values. In addition, we compare the estimates from our model under an all-case (AC) analysis – i.e., incorporating observations with missing values – and a complete-case (CC) analysis. (Chen et al. 2008).

[26] We omit the results for Model A-1 since, as seen in 2.1, this model fails to converge for large values of $\bar{\pi}^{1|0}$ and $\bar{\pi}^{0|1}$.

Table 2: Posterior means and 95% credible intervals for β with missing data

Average misreport probabilities	Missing data patters	Estimator	β_0	β_1	β_2
-	-	True values	-1	1	1
$\bar{\pi}^{1\|0} \approx 12\%$[a] $\bar{\pi}^{0\|1} \approx 18\%$	only $\tilde{y}_i$: 14.1%[b] only $x_{i,2}$: 25.4% $\tilde{y}_i$ and $x_{i,2}$: 8.2%	Ignoring Misreporting	-0.98 (-1.15, -0.83)	0.55 (0.44, 0.67)	1.37 (1.12, 1.63)
		Proposed Method			
		AC	-0.92 (-1.43, -0.55)	0.82 (0.55, 1.24)	0.92 (0.34, 1.59)
		CC	-0.72 (-1.31, -0.05)	0.69 (0.20, 1.21)	0.67 (-0.15, 1.45)
		Model A-2	-2.33 (-3.60, -1.41)	1.61 (0.94, 2.46)	3.40 (2.14, 5.13)
$\bar{\pi}^{1\|0} \approx 8\%$[c] $\bar{\pi}^{0\|1} \approx 7\%$	$\tilde{y}_i$: 35.5%[d] $x_{i,2}$: 11.8% $\tilde{y}_i$ and $x_{i,2}$: 13.9%	Ignoring Misreporting	-0.85 (-1.06, -0.62)	0.72 (0.56, 0.89)	1.05 (0.72, 1.36)
		Proposed Method			
		AC	-0.94 (-1.43, -0.52)	0.95 (0.64, 1.38)	0.97 (0.39, 1.63)
		CC	-1.29 (-2.58, -0.61)	1.57 (0.87, 2.82)	1.30 (0.54, 2.51)
		Model A-2	-1.28 (-1.79, -1.89)	1.11 (0.80, 1.56)	1.64 (1.09, 2.35)

[a] $\pi_i^{1|0}|x_{i,2} = 0 : 0.13$, $\pi_i^{1|0}|x_{i,2} = 0 : 0.5$, $\pi_i^{0|1}|x_{i,2} = 0 : 0.47$, $\pi_i^{0|1}|x_{i,2} = 1 : 0.07$.
[b] $\phi_{1,0} = 1.2$, $\phi_{1,1} = 0.5$, $\phi_{1,2} = 0.9$, $\phi_{2,0} = 0.5$, $\phi_{2,1} = 0.5$
[c] $\pi_i^{1|0}|x_{i,2} = 0 : 0.03$, $\pi_i^{1|0}|x_{i,2} = 0 : 0.18$, $\pi_i^{0|1}|x_{i,2} = 0 : 0.11$, $\pi_i^{0|1}|x_{i,2} = 1 : 0.02$.
[d] $\phi_{1,0} = 0.7$, $\phi_{1,1} = 0.45$, $\phi_{1,2} = 1.1$, $\phi_{2,0} = 0.7$, $\phi_{2,1} = 0.4$

As in the case with no missing data, failing to account for misreporting leads to underestimate the coefficient for x_1, while β_2 is overestimated under Model A-2. In contrast, the point estimates from our model are always closer to the true parameter values, which lie comfortably within the estimated 95% credible intervals for the two sets of misclassification levels and missing patterns reported. Also, the comparison of the AC and CC analyses shows that list-wise deletion leads to less accurate and precise estimates when the missing data mechanism for $\tilde{y}_i$ and the misclassification structure are both highly correlated with the same covariate - as is the case of the two scenarios presented in Table 2. However, additional simulation results (not shown) suggest that this is not necessarily true when $m_i^{\tilde{y}}$ is strongly related to x_1 but not to x_2. In all cases, though, our method leads again to more accurate estimates of the covariate effects than any of the other estimation approaches considered.

3. AN EMPIRICAL APPLICATION: CORRECTING FOR MISREPORTING IN THE ANALYSIS OF VOTER TURNOUT

Next, we illustrate the potential consequences of misreporting in the context of estimating the determinants of voter turnout and provide three different applications of our methodology using data from all the validated ANES surveys between the 1978 and 1990.[27] This dataset comprises three Midterm (1978, 1986, 1990) and three Presidential elections (1980, 1984, 1988), and has the obvious advantage of allowing us to directly compare the estimates from our model to a known benchmark, i.e., the same model estimated directly on the validated vote. We assume the validated vote to be the "gold-standard" measure of turnout, although there is considerable disagreement on this point (Burden 2000; Mcdonald 2007). The concern is that the validation studies are far from perfect. As stated at the outset, vote validation is expensive and difficult. The ANES is conducted in two parts, a pre- and post- election survey. In the studies from 1978, 1980, 1984, 1986, 1988 and 1990 there were in total 11,632 completed post election surveys. Unfortunately of these completed surveys, the ANES was unable to validate 2,189 respondents, about 19.8 percent of the usable sample.[28] The majority of these failures were caused either because no registration records were found or because the local election office refused to cooperate with the ANES. If we are willing to maintain the assumption that these errors are essentially random (in the sense of being independent of the characteristics of interest), then there is no real harm done. The measurement error will merely result in less efficient estimates of the misreporting model and a corresponding reduction in efficiency of the corrected turnout model. However, if there is systematic error, then we are just substituting one form of measurement error for another.

In Section 3.1, we estimate a simple model of the determinants of the turnout decision using both *self-reported* and *validated* turnout as the dependent variable in order to assess the consequences of ignoring misreporting. In 3.2, we re-estimate the turnout model with self-reported vote but applying our proposed solution to correct for misreporting, using a random sample of each survey as a validation sub-study. In 3.3, we apply our correction for misreporting under an external validation design, using information from previous ANES studies to correct for misreporting in the

[27] We use data from the 1978–1990 validated studies in order to preserve the comparability of the survey questions regarding the conditions of the interview; we will use this information to model the conditional probability of misreporting. While we illustrate the application of our method analyzing ANES data in view of the fact that it is the most widely used survey for studying U.S. turnout (Burden 2000), the main substantive results reported in this Section hold for the Current Population Survey as well, and are available from the authors upon request.

[28] The rate of non-validation varies considerably across Election Studies, from around 2% of sample in 1978 to more than 31% in 1990.

main sample under analysis. Both applications are based on a complete-case analysis. We deal with the problem of incomplete data in 3.4, where we account for item and unit non-response using the model-based, fully Bayesian imputation approach described in Section 1.3.

3.1. *Turnout misreporting in the 1978–1990 ANES*

As mentioned in the Introduction, it has long been established in the political science literature that survey respondents often report to have voted when they did not actually do so (Ansolabehere and Hersh 2008; Bernstein, Chadha and Montjoy 2001; Clausen 1968; Katosh and Traugott 1981; Miller 1952; Parry and Crossley 1950; Sigelman 1982; Silver, Anderson and Abramson 1986). Figure 5 illustrates the differences between turnout rates computed from self-reported and validated vote in the six ANES studies under analysis. Validated turnout is systematically lower than reported turnout, and while both rates tend to follow similar trends, differences vary considerably across years, ranging from 7 percentage points in 1990 to more than 15 percentage points in 1980. The percentage of survey respondents who claimed to have voted but did not do so according to the validated data was 17.3 percent, and more than 28% of those who did not vote according to the official records responded affirmatively to the turnout question. In contrast, only 84 respondents in the 1978-1990 ANES studies reported not voting when the official record suggested they did, representing 0.7% of the sample respondents. Additional descriptive statistics on vote misreporting in the 1978–1990 validated ANES can be found in Table 9 in Appendix B.

In order to examine whether such high rates of overreporting affect inferences on the determinants of the turnout decision, we fit two hierarchical probit models allowing for election year and regional effects with both *self-reported* and *validated turnout* as the response variable:

$$\Pr[\widetilde{y}_i = y_i^{Reported}] \sim Bernoulli(p_i);$$
$$p_i = \Phi(\lambda_t + \eta_r + \beta\,'\mathbf{x}_i)$$

and

$$\Pr[y_i = y_i^{Validated}] \sim Bernoulli(p_i);$$
$$p_i = \Phi(\lambda_t + \eta_r + \beta\,'\mathbf{x}_i);$$

where the k=1,...,K elements of β are assigned diffuse prior distributions:

$$\beta_k \sim N(\mu_{\beta_k}, \sigma^2_{\beta_k})$$

and λ_t and η_r are election- and region- random effects distributed

$$\lambda_t \sim N(\mu_\lambda, \sigma^2_\lambda), \;\; t = 1978, 1980, 1984, 1986, 1988, 1990;$$
$$\eta_r \sim N(\mu_\eta, \sigma^2_\eta), \;\; r = \text{Northeast, North Central, South, West}$$

The regressors included in $\mathbf{x}_i$ are indicators for demographic and socio-economic conditions and political attitudes: *Age*, *Church Attendance*, *Education*, *Female*, *Home owner*, *Income*, *Nonwhite*, *Party Identification* and *Partisan Strength*. A description of the coding used for each of the variables may be found in Appendix B.1. We should note that, while this specification includes some of the variables most commonly used in models of voter turnout found in the literature (Ansolabehere

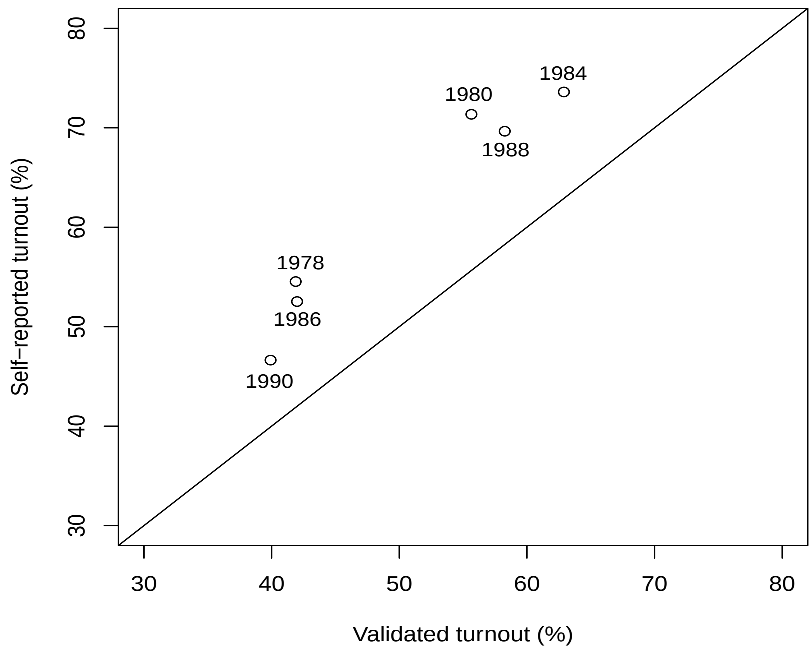

Figure 5: ***Estimated Turnout from Self-reported vs. Validated Responses, 1978–1990.*** *The graph shows the self-reported and validated turnout from the 1978–1990 ANES only in years for which there were vote validation studies. Reported turnout rates are systematically larger than the validates ones.*

and Hersh 2008; Bernstein, Chadha and Montjoy 2001; Highton 2004; Leighley and Nagler 1984; Wolfinger and Rosenstone 1980), it does not examine the effect of other factors we might plausibly believe could alter turnout, such as political information (Alvarez 1997) or differences in state-level ballot laws (Wolfinger and Rosenstone 1980). The sample used in the analysis consists of 6,411 observations for the 6 elections under study and were constructed so that they are identical for both models. Only the respondents with no missing response or covariate values are included in the analysis; the remaining observations were dropped using list-wise deletion.

Figure 6 presents the main results fro both models.[29] The left panel summarizes the posterior distribution of the model's coefficients using self-reported vote as the dependent variable, and the right panel re-does the analysis with the ANES validated vote. Most of the parameter estimates are quite similar in both models, and inferences on the role of these predictors on the probability of voting agree with common expectations. For example, for both sets of estimates, older, wealthier and more educated respondents are more likely to turn out to vote. Also, strong partisans are on average 15 percentage points more likely to vote than independents, while respondents who attend church every week are 0.12 more likely to turn out to vote than those who never attend. Likewise, respondents are much more likely to turn out to vote in Presidential than in Midterm elections,

[29] Three parallel chains with dispersed initial values reached approximate convergence after 50,000 iterations, with a burn-in period of 5,000 iterations. In order to ensure that inferences are data dependent, several alternative values for the hyperparameters were tried, yielding essentially similar results.

and are less likely to vote if they live in the South. These results are essentially similar using either reported or validated vote as the dependent variable. However, there are some interesting differences between the two sets of results regarding the role of some socio-demographic variables such as gender and race. In particular, the mean posterior of the coefficient for the race indicator is more than twice as large (in absolute value) using validated vote than using self-reported vote as the dependent variable.

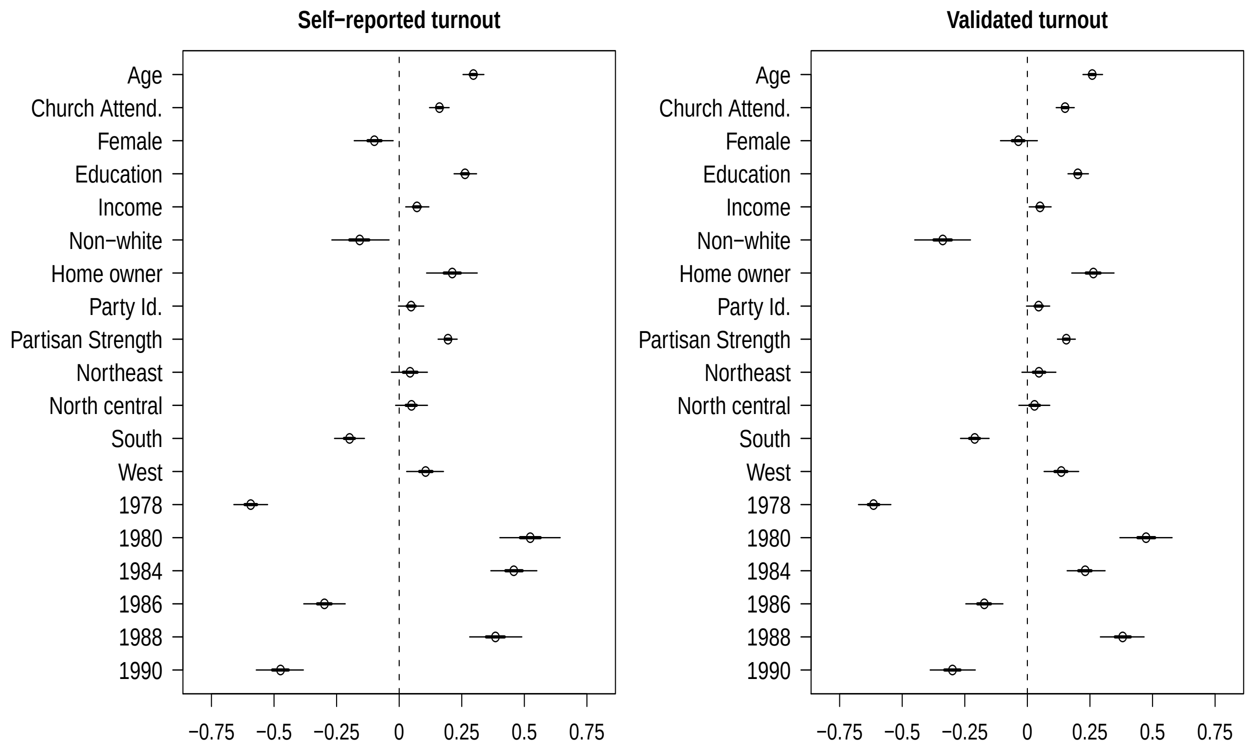

Figure 6: ***Coefficients of the probit models for Self-reported vs. Validated Turnout.*** *The graph summarizes the posterior distribution of the coefficients of the turnout model, using self-reported and validated vote as the response variable. The center dots correspond to the posterior means, the thicker lines to the 50% credible intervals, and the thinner lines to the 95% credible intervals.*

These differences in the parameter estimates can affect inferences drawn from both models regarding the impact of the covariates on the turnout decision. In order to illustrate this fact, Figure 7 plots the marginal effect of race on the probability of voting using reported and validated vote for each election under analysis. As seen in the figure, the negative effect of being *Non-white* on turnout is higher when validated vote is used as the response variable for each of the surveys considered: the average marginal effects (posterior means) are more than 6 percentage points higher than if we look only at the reported vote, with differences ranging from about 3 percentage points in the 1984 and 1986 elections to almost 11 points in the 1978 and 1988 elections. While a researcher using reported turnout would conclude that race had no significant effect on the probability of voting in the 1978 and 1988 elections at the usual confidence levels, the results obtained using validated

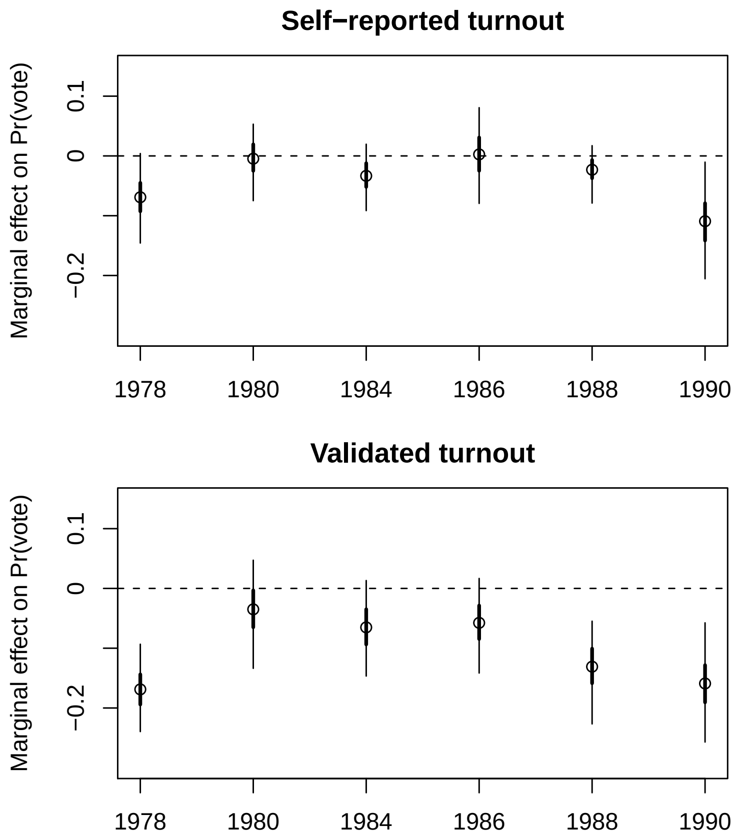

Figure 7: ***Marginal effect of race on turnout.*** *The graph shows the marginal effect of the race indicator on the likelihood of voting for each election year under study, using both reported and validated vote. The center dots correspond to the point estimates (posterior means), the thicker lines to the 50% credible intervals, and the thinner lines to the 95% credible intervals.*

data indicate otherwise.[30] Fitting a model of turnout using reported vote as the dependent variable will therefore tend to overpredict the probability of voting among non-white respondents and might in some cases affect substantive conclusions about the effect of race on turnout.

Finally, we examine whether over-reporting varies systematically with respondents' characteristics, fitting a probit model for $\Pr[\tilde{y}_i = 1|y_i = 0]$. As with the turnout model, the misreport model is fairly simple. The predictors include four variables that have been shown to be strongly correlated with overreporting in previous studies: *Age*, *Church Attendance*, *Education*, *Non-white*, and *Partisan Strength* (Ansolabehere and Hersh 2008; Belli, Traugott and Beckman 2001; Bernstein, Chadha and Montjoy 2001; Cassel 2003). In addition, we also include three additional covariates aimed at capturing some of the conditions of the interview. The first is an indicator of whether the interview was conducted while the respondent was alone. According to the "social pressures" argument (Cahalan 1968; Loftus 1975), a respondent should be more likely to lie about voting if others will learn of the statement. The other two variables are the interviewers' assessments of the

[30]In the case of the 1988 election, the marginal effect of Non-white estimated from the self-reported vote is not significant even at the 0.1 level

respondents' cooperation and sincerity during the interview.[31] Point and interval summaries of the posterior distribution of the model's parameters are presented in Figure 8. In line with previous analyses, we find that overreporters tend to be more educated, older, more partisan, and are more likely to be regular church attendees. Also, consistent with the results reported in Figures 6 and 7, being nonwhite has a positive effect on the probability of misreporting vote status: non-whites are on average 0.05 more likely to overreport than their white counterparts, and this effect is significant at the 0.1 level. Several scholars have argued that African Americans and Latinos feel pressured to appear to have voted due to the struggles and sacrifices needed to gain voting rights for their racial or ethnic group (Abramson and Claggett 1984; Belli, Traugott and Beckman 2001; Hill and Hurley 1984), although recent research has suggested that the relationship between race and overreporting is much more complex than previously thought and depends on the demographic and geographical context (Ansolabehere and Hersh 2008; Bernstein, Chadha and Montjoy 2001; Fullerton, Dixon and Borch 2007).[32] None of the other variables has a statistically significant effect on misreporting at the usual confidence levels. In particular, the interviewers seem unable to pick up a "feeling" that is not otherwise captured by the characteristics observable from the survey. This is probably caused by the fact that very few of the interviewers were willing to rank a respondent as uncooperative and/or insincere.[33]

Hence, the results from these simple models indicate that the probability of misreporting varies systematically with characteristics we might be interested in, and that failing to account for misreporting may affect parameter estimates and inferences about the determinants of voter turnout drawn from non-validated survey data. Unfortunately, as mentioned in the Introduction, the ANES has stopped conducting validation studies due to the cost and difficulty in collecting the data as well as to the fact that few researchers used the validated data. The next three sections allow us to evaluate the performance of our proposed method to correct for misreporting and improve estimates and inference obtained from self-reported turnout. Although our model accounts for the possibility of two types of misreporting, we saw before that virtually no one reports not voting when they did, and thus $\pi_i^{0|1}$ would be poorly estimated (Prescott and Garthwaite 2005). Therefore, in the applications below we will assume that $\pi_i^{0|1} = 0$, and we therefore only need to account for $\pi_i^{1|0}$.

3.2. *Correcting for misreporting using a validation sub-sample*

We first apply our method assuming an internal validation design. As in the simulation exercise in Section 2.1, we randomly assign half of the respondents in each of the 1978–1990 surveys to be the validation sub-study and ignore the validated data for the remaining respondents. We then used the information from the validated sub-sample to correct for over-reporting in the main sample, equally weighting both datasets. For illustrative purposes, we fit the same turnout and misreport

[31] All interviewers in the 1978 – 1990 ANES were asked to rate the level of cooperation and sincerity of the respondent after the completion of the survey.

[32] It is worth mentioning that this relationship between race and vote over-reporting could also be associated to the socio-economic status of the non-white population. If it is the case that nonwhites, who are more concentrated in poorer areas, are more likely to be incorrectly validated or excluded from the validation studies because no records can be found (e.g., due to poorly staffed and maintained election offices), then this result - as well as those reported in Figures 6 and 7 - could very well be an artifact. While it is difficult to rule this claim out, addressing this concern is beyond the focus of this paper. Hence, as noted above, we proceed as if the validated data provides "gold-standard" information on turnout, or is at least not subject to systematic bias.

[33] Only 1.3% of all the respondents in the sample were ranked as uncooperative by the ANES interwievers, and only 0.7% were deemed to be "often insincere".

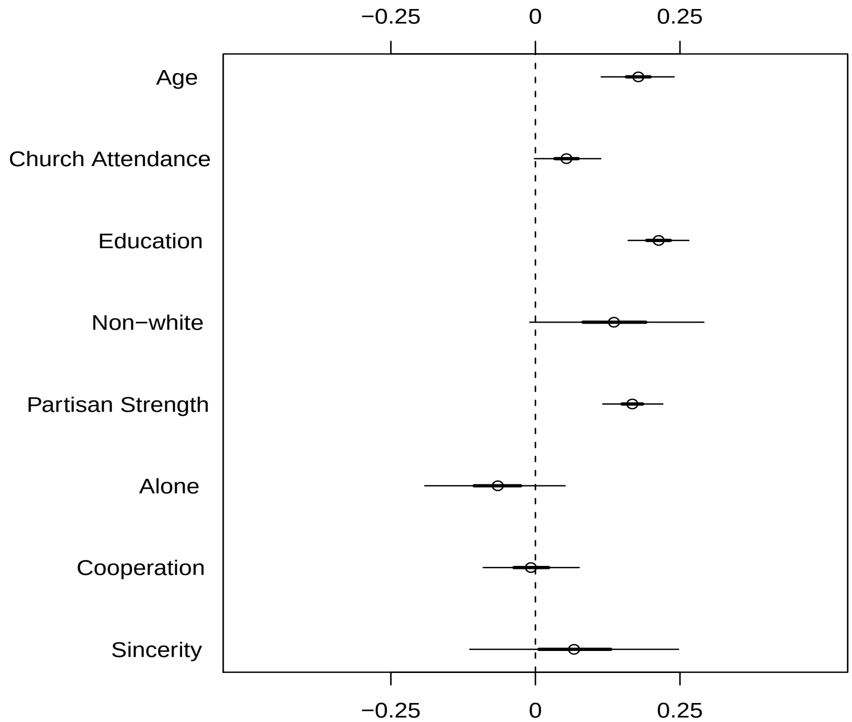

Figure 8: ***Determinants of misreporting.*** *The graph shows the parameter estimates for the model of over-reporting. The center dots correspond to the point estimates (posterior means), the thicker lines to the 50% credible intervals, and the thinner lines to the 95% credible intervals.*

models described in 3.1 for all the ANES studies considered. Nonetheless, as indicated above, the probability of voting is considerably higher in Presidential than in Midterm elections, and it is likely that different factors affect turnout in different election years. More importantly, the patterns of overreporting have also been shown to differ substantially across types of races and election years (Cassel 2003). As a result, the misreport model does not predict over-reporting very well: the mean error rate of the misreport model across election studies is 36%, while a null model that simply predicts that no respondent overreports has an error rate of 31%. The model correctly classifies 64% of the survey respondents in cases, and the mean predicted probability of misreporting averaged across simulations is 0.45; ideally this would be near zero or one for the entire sample (Gelman and Hill 2007). Therefore, while the simulation results from Section 2 suggest that our approach is quite robust to misspecification of the model of misreporting, we note that the performance of our proposed method would benefit from better modeling of the misreport process.

Figure 9 summarizes the posterior distribution of the coefficients of selected regressors estimated using validated, self-reported vote, and corrected self-reports for the two ANES studies with lowest (1978) and largest (1984) percentage of overreporters (See Table 9 in Appendix B). Assuming that the parameters estimated using validated vote are the "correct" estimates, the point estimates (posterior means) from our model for the two elections are between 32% and 92% closer to the "true" values of each of the parameters than the estimates ignoring overreporting. In addition, like the "true" estimate, the estimate of $\beta_{Non-white}$ under our approach is significantly negative at the 0.05 level for the 1978 ANES.

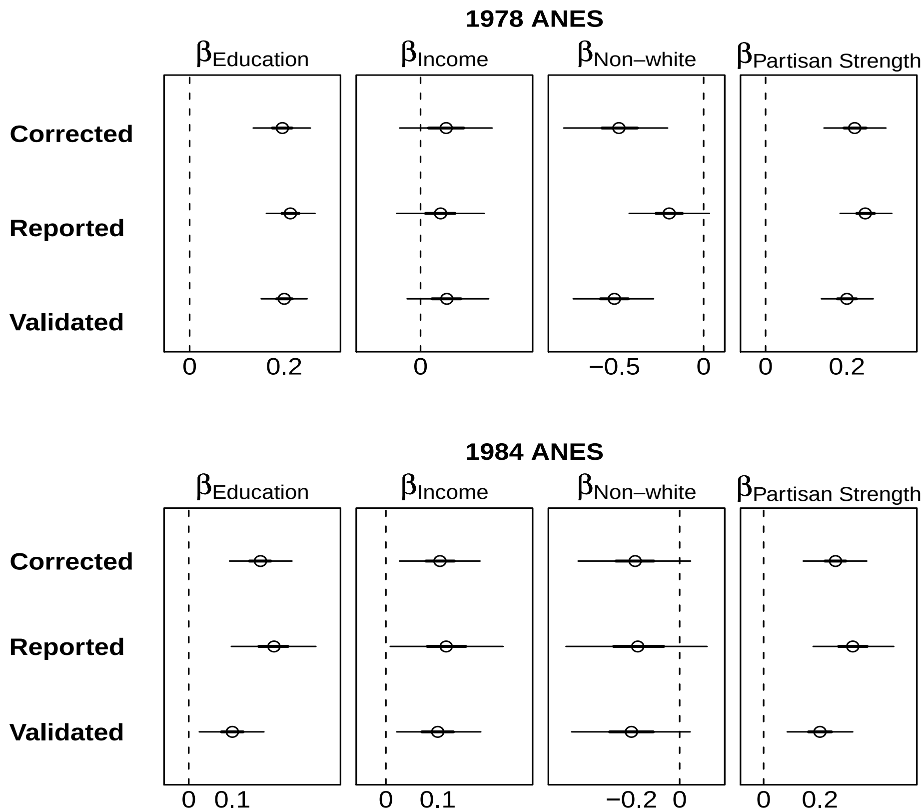

Figure 9: ***Posterior summaries for selected parameters under an internal validation design*** *The figure plots point and interval summaries of the posterior distributions of selected coefficients for the 1978 and 1984 ANES Presidential elections, using corrected, self-reported, and validated vote. The center dots correspond to the posterior means, the thick horizontal lines to the central 50% credible intervals, and the thin lines to the central 95% credible intervals from the three different models.*

Figure 10, in turn, plots the marginal effect of race on the probability of voting estimated using our approach to correct for misreporting. A comparison of the results in the left panel of the figure with those presented in Figure 7 above shows that, after correcting for misreporting, the impact of race in the 1978 and 1988 elections is now statistically significant at the usual confidence levels. Moreover, as seen in the right panel, the point estimates from our model are closer to the "true" effects than those estimated from the model using self-reported vote for all the ANES studies, with differences ranging between 1 and 9 percentage points. Therefore, the evidence presented in this Section indicates that, even with the very simple model of misreporting estimated here, the improvements in the accuracy of the parameter estimates obtained using our method are important, and can eventually change the substantive conclusions drawn regarding the effect of relevant covariates on the turnout decision.

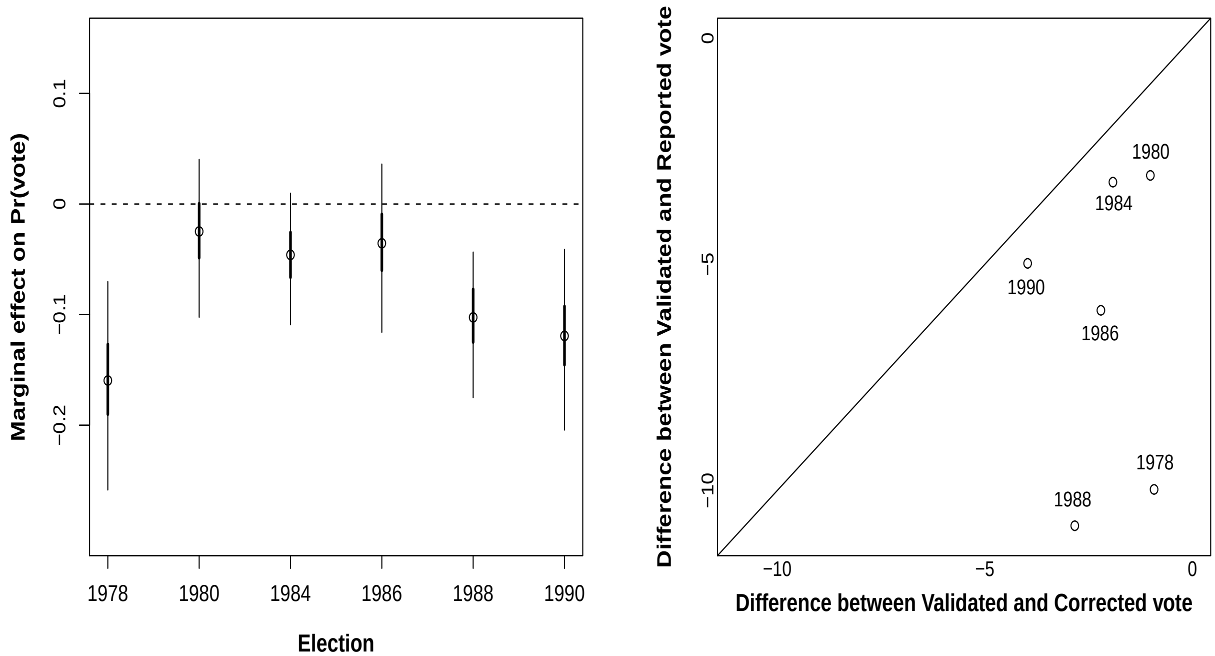

Figure 10: ***Marginal effect of race on turnout estimated under our proposed method.*** *The left panel of the graph plots the point and interval (50% and 95%) estimates of the marginal effect of race on the probability of voting estimated from our model to correct for misreporting. The right panel compares the point estimates from our model and the model ignoring misreporting with the estimates obtained using the validated data.*

3.3. *Correcting for misreporting under an external validation design*

We also apply our correction for misreporting assuming an external validation design, ignoring the validated vote for the sample under analysis and incorporating information on the misreport probabilities and regression parameters from other ANES studies. Figure 11 illustrates the results of this exercise, plotting the marginal posterior distribution of selected coefficients for the 1988 and 1992 Presidential elections obtained by updating the corresponding posteriors from previous validated ANES surveys.

The upper panel the compares the posterior distributions of $\beta_{Education}$, β_{Income}, $\beta_{Non-white}$ and $\beta_{Partisan\ Strength}$ for the 1988 ANES, the last Presidential election for which vote validation is available, using validated, self-reported and corrected vote. In order to implement our correction for misreporting, we used auxiliary data from the two previous Presidential elections for which validated turnout data was collected (1980 and 1984). As seen in the figure, the marginal posterior means and modes from the model accounting for overreporting are in all cases closer to "true" values than those obtained from the unadjusted self-reports. Again, as the "correct" estimate, the estimate of $\beta_{Non-white}$ under our model is significantly negative at the 0.05 level. In the case of the 1992 ANES, for which there is no validated data, we implemented our correction for misreporting using information from the previous presidential elections for which vote validation was conducted

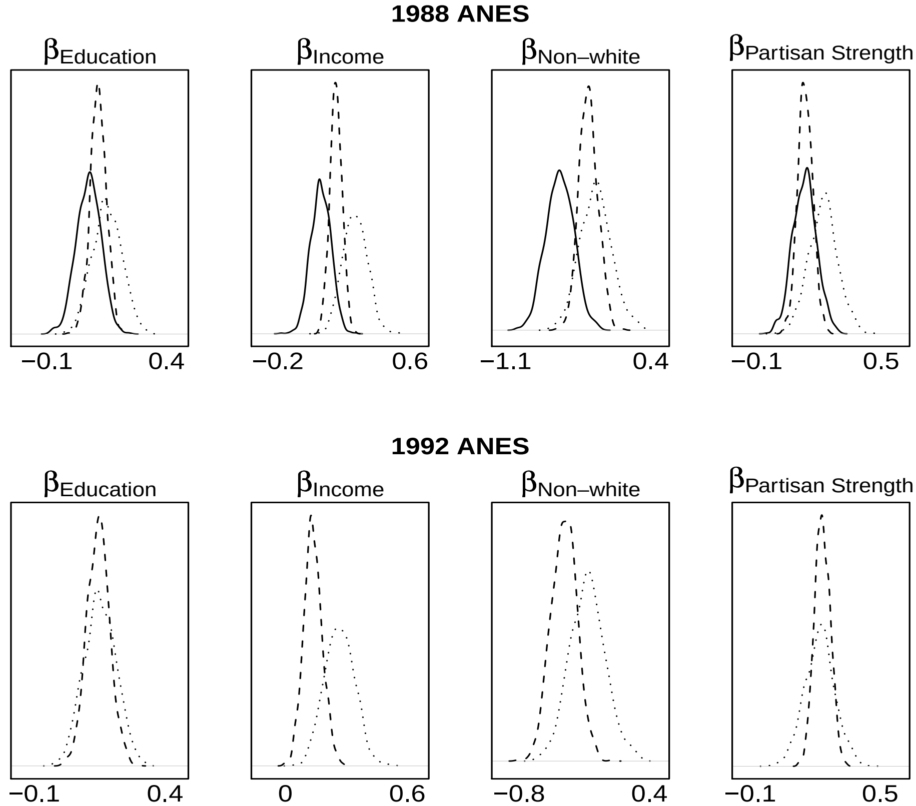

Figure 11: ***Posterior densities of β under an external validation design.*** *The figure compares the posterior densities of selected coefficients for the 1988 and 1992 Presidential elections. The solid lines plot the posterior distributions of the parameters estimated from the validated vote, the dotted lines represent the estimates obtained using self-reported vote, and the dashed lines the ones obtained adjusting for misreporting.*

(1980, 1984 and 1988) and compared the estimates from our model with those from a model using self-reported vote. As seen in the lower panel of Figure 11, the posterior distribution of some of the parameters - $\beta_{Education}$, $\beta PartisanStrength$ - remain essentially unchanged when applying the correction for misreporting. However, using auxiliary information does affect the posterior distribution of the coefficients of *Income* and *Non-white*. In particular, accounting for misreporting substantially affects the marginal posterior distribution of $\beta_{Non-white}$: the mean posterior is more than twice as large (in absolute value) when using the corrected self-reports, and the effect of *Non-white* on the probability of turning out to vote is significantly negative at the 0.05 level, while it is not significant even at the 0.2 level when estimated using self-reported vote. Similar results hold when applying our model to correct for misreporting in the 1994 ANES - for which, again, vote validation was not conducted - using validated turnout data from previous Midterm elections.

We also conducted a series of sensitivity analyses aimed at assessing the robustness of the parameter estimates to changes in the composition of the auxiliary data used to correct for misreporting and in the weight assigned to the validated *vis a vis* the main sample. Figure 12 summarizes some of the results for the 1988 and 1992 ANES. The left panel plots point and interval summaries for

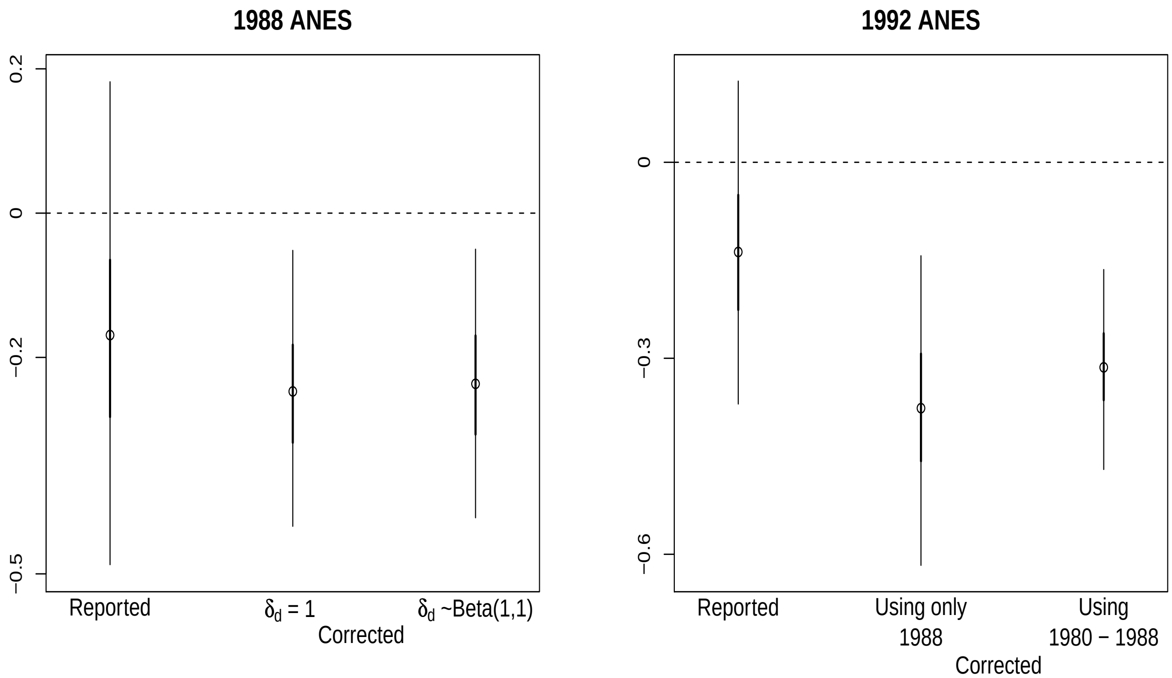

Figure 12: ***Sensitivity analysis for the external validation design.*** *The graph summarizes the posterior distribution of $\beta_{Non-white}$ from our model for the 1988 and 1992 elections, using alternative strategies to incorporate information from previous validated ANES studies. The estimates are compared to those obtained using self-reported vote. The center dots correspond to the posterior means, the thicker lines to the 50% credible intervals, and the thinner lines to the 95% credible intervals.*

$\beta_{Non-white}$ from our model for the 1988 ANES using two different sets of values for the weighting parameters δ_d in Equation 9: a point mass prior $\delta_d = 1$ with probability 1 $\forall d$, and uniform $Beta(1,1)$ priors $\forall d$, where $d = 1980, 1984$. In the first case, the validated and main samples are pooled together and the estimates of β for the main sample are obtained by updating the posteriors from the previous ANES surveys *via* Bayes' theorem. In the second case, we allow for different *a posteriori* weights for each of the validated samples, thus accommodating heterogeneity between the previous ANES studies. The right panel, in turn, compares the estimates from our model for the 1992 for the cases in which only validated data from the immediate previous (1988) or from all the previous (1980, 1984, 1988) Presidential elections is used to adjust for misreporting.[34] For both election years, the estimates from our model are compared to those from the unadjusted self-reports.

As illustrated in the figure, the posterior standard deviations of β tend to decrease with the amount of auxiliary data used to correct for misreporting in the main sample, but the point estimates (posterior means) and the main substantive conclusions about β seem to be quite robust to changes in the values of δ and in the size and heterogeneity of the auxiliary data. In particular,

[34] For the 1992 ANES, we fixed the value of δ at 1 for this sensitivity analysis.

correcting for overreporting using information from previous validated studies leads to stronger negative effects of being *Non-white* on the probability of voting than using self-reported vote, with differences of approximately 4 and 9 and percentage points for the 1988 and 1992 ANES, respectively.

3.4. *Accounting for item and unit non-response*

Both applications of our methodology in Sections 3.2 and 3.3 have been based on a complete-case analysis, including in the sample only those respondents for whom both the response to the turnout question and all the relevant covariates are completely observed. When respondents with missing covariates differ systematically from those with complete data with respect to the outcome of interest, this approach may lead to significantly biased parameters and inference (Little and Rubin 2002). In our sample from the 1978–1990 ANES studies, 14.5% of whites and 20.9% of non-whites have missing covariate values (other than race), and the percentage of missingnes for the self-reported vote is almost 1.8 times larger for the latter. Since the evidence above indicates that voting patterns vary systematically with race, inferences from a complete-case analysis may be quite misleading in this setting (Ibrahim et al. 2005). In addition, list-wise deletion due to missing values in the response variable and/or the predictors leads to discard almost 45% of the respondents in the 1978–1990 ANES and more than two-thirds of the respondents in the 1994 ANES, so that complete-case analyses are extremely wasteful and potentially inefficient. Table 10 in Appendix B reports the rates of item nonresponse for all the variables included in the turnout models from Sections 3.2 and 3.3.

In order to accommodate item and unit non-response, we implement the approach described in Section 1.3, fitting a separate model for each of the ANES studies.[35] Based on Equation 12, we specified probit regression models for all the dichotomous covariates in the model – *Female*, *Non-white*, *Own Home*, and *Alone* –, while the remaining categorical covariates were assigned conditional normal distributions and discrete values were afterwards imputed for the missing responses (Lipsitz and Ibrahim 1996; Gelman, King and Liu 1998).[36] In all cases, we assigned vague independent normal priors for the components of α.

Figure 3.4 illustrates the results for the 1978 and 1992 ANES. For the former, 31% of the survey respondents have at least 1 missing covariate value, and 0.5% of the respondents failed to answer the turnout question, while the corresponding rates for the latter are 47% and 9%, respectively. A complete-case analysis would keep 77% of our sample for the 1978 ANES, and only 42% for the 1992 ANES. The left panel of the figure summarizes the marginal posterior distribution of $\beta_{Non-white}$ for the 1978 ANES using reported, validated and corrected vote. As in Section 3.2, our correction for misreporting was implemented based on auxiliary information from a random sub-sample of the ANES survey. The right panel plots the estimates for the 1992 ANES, for which we use validated turnout data from the previous Presidential elections, as in Section 3.3. In both cases, estimates obtained using Bayesian imputation are compared to those from the complete-case analyses.

Two interesting facts emerge from the figure. First, for both election-studies, the marginal posterior distribution for $\beta_{Non-white}$ estimated using our the Bayesian imputation model is not statistically different from the that obtained using list-wise deletion, at least at the 0.05 level.

[35] See Gelman, King and Liu (1998) for an approach to multiple imputation for multiple surveys using hierarchical modeling.

[36] The substantive results are essentially unchanged if, instead of the normal distributions, one-dimensional conditional gamma distributions are specified for these covariates, all of which are strictly positive.

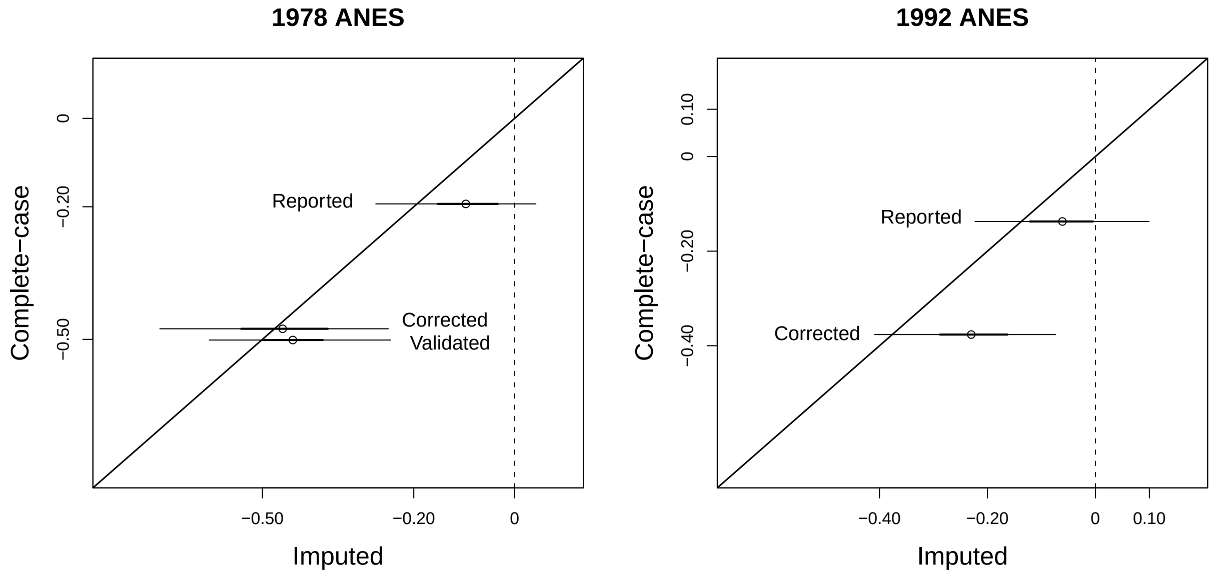

Figure 13: ***Posterior summaries for*** $\beta_{Non-white}$ ***with list-wise deletion versus Bayesian imputation.*** *The graph plots point and interval summaries for* $\beta_{Non-white}$ *for the 1978 and 1992 ANES, using list-wise deletion and fully Bayesian imputation. The center dots correspond to the point estimates (posterior means), and the horizontal bars indicate the 90% and 50% confidence intervals for the models with imputed missing values.*

However, the standard errors tend to be lower when missing values are imputed than under list-wise deletion. This result holds in fact for most of the election-years under analysis, suggesting that by omitting the cases with missing values, much information is lost on the variables that are completely or almost completely observed, thus leading to less efficient parameter estimates (Ibrahim, Chen and Lipsitz 2002; Ibrahim et al. 2005). This is likely to be an important concern in the Election Studies examined here, given that there is substantial variation in the rates of item nonresponse, with most of the variables exhibiting relatively low percentage of missing values while a few others show very high rates of nonresponse (see Appendix B). Second, imputing missing values does not change the substantive findings reported above regarding the performance of our methodology. The results for the 1978 ANES show that the estimated effects from our model correcting for misreporting are again closer to the benchmark case – using validated vote– than the effects estimated using recalled vote, and this result holds for all the ANES with validated vote. For the 1992 election, the marginal effect of race obtained from the corrected turnout model is also higher than in the uncorrected model, as was in the complete-case analysis. For both elections, once again, the main substantive conclusions regarding the effect of being *Non-white* on the probability of voting drawn from the model correcting for misreporting differ from those obtained using recalled vote.

4. CONCLUDING REMARKS

Survey data are usually subject to measurement errors, generally referred to as classification errors when affecting discrete variables. In the political science literature, misclassification of binary dependent variables has received considerable attention in the context of estimating the determinants of voter turnout. High rates of overreporting have been documented in survey instruments commonly used to study turnout in the U.S., such as the American National Election Study (ANES) and the Current Population Survey (CPS), and most previous research has found that misreporting varies systematically with some of the relevant characteristics affecting the turnout decision.

In the presence of misreporting, standard binary choice models will generally yield biased parameter estimates and inaccurate standard errors and may lead to erroneous substantive conclusions. This paper develops a simple Bayesian method to correct for misreporting using information on the misreport mechanism from auxiliary data sources. Our model does not require full validation studies to be conducted every time a researcher is concerned about potential misreporting. As long as enough data exists to reasonably estimate the misreporting probabilities, our approach can be applied for drawing inference from the non-validated samples, improving the accuracy of the parameter estimates and inferences on the effect of covariates of interest on the true response *vis a vis* standard models ignoring misclassification and methods assuming constant misreport rates. This is clearly important, since obtaining "gold-standard" data is usually quite expensive and time consuming, and thus restricting the analysis only to validated studies will generally lead to discard large amounts of useful information, as in the case of the ANES.

The proposed model is fully general and modular, can be easily implemented using freely available software, and can be readily applied in the case of missing data in the response and/or covariates. While we illustrate our technique using turnout data from the American National Election Study, it could be applied in general to account for potential misclassification of a binary dependent variable in many other situations in which auxiliary data on the misreport structure is available (Bound, Brown and Mathiowetz 2001; Molinari 2003). Extensions to more general discrete choice models are also straightforward. Potential avenues for future research would be to use semi- or non-parametric methods to estimate both the misreporting and turnout models (Horowitz and Manski 1995; Molinari 2003), simultaneously account for response and covariate measurement errors within our model (McGlothlin, Stamey and Seaman 2008), and explore the possibility of incorporating semi-parametric approaches for inference with missing data (Chen and Ibrahim 2006; Robins and Rotnitzky 1995; Rotnitzky and Robins 1995).

While the primary focus of the paper has been on estimation techniques as opposed to substantive findings, the empirical application of our model to the analysis of the determinants of voter turnout has clear implications for researchers interested in race. Our results confirm that race does have a clear negative impact on turnout, and suggest that the null previous findings have been probably due to problems of misreporting, as had been argued by Abramson and Claggett (1984, 1986*a*, 1991). With the correction for misreporting developed in this paper, researchers could now better estimate the effect of race over the length of the ANES datasets and not just for the few years with validated turnout data. In addition, researchers might wish to revisit Wolfinger and Rosenstone (1980) findings of the effect of registration laws to see if properly correct misreporting re-enforces or diminishes their findings.

Table 3: Misreport probabilities under Experimental Design B

Average misreport rates, $\bar{\pi}^{1\|0}, \bar{\pi}^{0\|1}$	$\pi_i^{1\|0}$		$\pi_i^{0\|1}$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
2%	0.01	0.12	0.06	0.008
5%	0.03	0.18	0.11	0.02
10%	0.05	0.35	0.24	0.04
20%	0.13	0.50	0.47	0.07

Table 4: Posterior means and 95% credible intervals - Design A

$\pi^{1\|0}, \pi^{0\|1}$	Estimator	β_0	β_1	β_2
-	True values	-1	1	1
0.02	Ignoring Misreporting	-0.86	0.88	0.92
		(-1.01, -0.74)	(0.77, 0.99)	(0.74, 1.11)
	Proposed method	-0.99	0.99	1.04
		(-1.19, -0.82)	(0.83, 1.17)	(0.80, 1.27)
	Model A-1	-1.16	1.15	1.20
		(-1.56, -0.86)	(0.89, 1.48)	(0.89, 1.61)
	Model A-2	-1.00	1.01	1.05
		(-1.20, -0.82)	(0.85, 1.18)	(0.84, 1.30)
0.05	Ignoring Misreporting	-0.82	0.78	0.90
		(-0.96, -0.68)	(0.68, 0.89)	(0.71, 1.08)
	Proposed method	-1.00	1.02	1.10
		(-1.21, -0.80)	(0.84, 1.22)	(0.84, 1.38)
	Model A-1	-1.05	1.08	1.20
		(-1.41, -0.76)	(0.82, 1.41)	(0.89, 1.62)
	Model A-2	-1.03	1.04	1.16
		(-1.26, -0.84)	(0.87, 1.24)	(0.90, 1.44)
0.10	Ignoring Misreporting	-0.67	0.64	0.64
		(-0.79, -0.55)	(0.55, 0.74)	(0.48, 0.80)
	Proposed method	-0.96	0.88	0.86
		(-1.26, -0.70)	(0.69, 1.10)	(0.55, 1.21)
	Model A-1	-0.81	0.77	0.76
		(-1.32, -0.45)	(0.56, 1.15)	(0.54, 1.26)
	Model A-2	-1.03	0.96	0.94
		(-1.34, -0.78)	(0.76, 1.20)	(0.66, 1.26)

Continued on next page

Table 4 – continued from previous page

$\bar{\pi}^{1\|0}, \bar{\pi}^{0\|1}$	Estimator	β_0	β_1	β_2
0.20	Ignoring Misreporting	-0.46	0.51	0.50
		(-0.58, -0.34)	(0.42, 0.60)	(0.34, 0.66)
	Proposed method	-1.01	1.00	0.97
		(-1.47, -0.62)	(0.68, 1.42)	(0.48, 1.52)
	Model A-1	-0.71	0.75	0.74
		(-1.21, -0.34)	(0.52, 1.12)	(0.44, 1.16)
	Model A-2	-0.98	1.07	1.02
		(-1.38, -0.68)	(0.79, 1.39)	(0.64, 1.46)

Constant misclassification rates.
Sample size: N, M = 1,000.

Table 5: Posterior means and 95% credible intervals - Design B

$\bar{\pi}^{1\|0}, \bar{\pi}^{0\|1}$	Estimator	β_0	β_1	β_2
-	True values	-1	1	1
0.02	Ignoring Misreporting	-0.95	0.93	1.04
		(-1.10, -0.82)	(0.81, 1.05)	(0.85, 1.23)
	Proposed method	-0.94	0.99	0.99
		(-1.12, -0.77)	(0.85, 1.16)	(0.77, 1.22)
	Model A-1	-1.18	1.12	1.23
		(-1.55, -0.89)	(0.90, 1.37)	(0.94, 1.56)
	Model A-2	-1.02	1.01	1.13
		(-1.20, -0.85)	(0.87, 1.16)	(0.91, 1.34)
0.05	Ignoring Misreporting	-0.92	0.76	1.05
		(-1.06, -0.79)	(0.65, 0.86)	(0.87, 1.23)
	Proposed method	-1.01	1.01	1.02
		(-1.23, -0.80)	(0.82, 1.21)	(0.75, 1.29)
	Model A-1	-	-	-
	Model A-2	-1.20	0.99	1.33
		(-1.44, -0.97)	(0.81, 1.19)	(1.07, 1.64)
0.10	Ignoring Misreporting	-0.88	0.70	1.18
		(-1.01, -0.75)	(0.59, 0.80)	(0.98, 1.37)
	Proposed method	-0.95	1.04	1.07
		(-1.24, -0.67)	(0.82, 1.28)	(0.75, 1.43)
	Model A-1	-	-	-
	Model A-2	-1.34	1.07	1.73
		(-1.66, -1.08)	(0.85, 1.32)	(1.40, 2.13)

Continued on next page

Table 5 – continued from previous page

$\bar{\pi}^{1\|0}, \bar{\pi}^{0\|1}$	Estimator	β_0	β_1	β_2
0.20	Ignoring Misreporting	-1.16	0.45	1.77
		(-1.30, -1.01)	(0.36, 0.55)	(1.58, 1.99)
	Proposed method	-1.06	0.98	1.27
		(-1.55, -0.56)	(0.66, 1.36)	(0.70, 1.82)
	Model A-1	-	-	-
	Model A-2	-2.58	1.00	3.76
		(-3.55, -1.92)	(0.70, 1.42)	(2.96, 4.92)

Covariate-dependent misclassification.
Sample size: N, M = 1,000.

Table 6: Posterior means and central 95% credible intervals for $\bar{\pi}^{1|0}$ and $\bar{\pi}^{0|1a}$

Monte Carlo design	True sample values	Proposed method	Model A-1	Model A-2
A	$\bar{\pi}^{1\|0} = 2.75$	2.85 (1.79, 4.29)	2.89 (1.77, 4.27)	6.06 (1.33, 11.28)
	$\bar{\pi}^{0\|1} = 1.84$	2.46 (1.20, 4.07)	2.39 (1.10, 4.07)	4.95 (0.29, 12.12)
	$\bar{\pi}^{1\|0} = 5.17$	5.29 (3.84, 7.02)	5.26 (3.77, 7.00)	5.68 (1.16, 11.12)
	$\bar{\pi}^{0\|1} = 4.47$	6.56 (4.40, 9.21)	6.56 (4.34, 9.25)	8.24 (1.11, 16.49)
	$\bar{\pi}^{1\|0} = 9.53$	10.35 (8.27, 12.71)	10.36 (8.08, 12.65)	6.74 (1.88, 11.72)
	$\bar{\pi}^{0\|1} = 9.13$	9.34 (6.79, 12.27)	8.87 (6.36, 11.82)	5.17 (0.74, 9.60)
	$\bar{\pi}^{1\|0} = 20.68$	21.23 (18.45, 24.28)	20.48 (17.57, 23.33)	11.92 (1.58, 21.77)
	$\bar{\pi}^{0\|1} = 20.26$	20.44 (16.47, 24.53)	20.55 (16.81, 24.44)	10.29 (0.45, 24.85)
B	$\bar{\pi}^{1\|0}$=2.42	1.62 (0.84, 2.71)	1.52 (0.70, 2.63)	4.55 (0.64, 9.73)
	$\bar{\pi}^{0\|1} = 2.37$	2.60 (1.23, 4.47)	2.20 (1.06, 3.75)	2.76 (0.13, 8.61)
	$\bar{\pi}^{1\|0} = 5.01$	5.14 (3.72, 6.97)	5.36 (3.79, 7.10)	-
	$\bar{\pi}^{0\|1} = 6.32$	5.01 (3.15, 7.36)	4.56 (2.71, 6.74)	-

Continued on next page

Table 6 – continued from previous page

Monte Carlo design	True sample values	Proposed method	Model A-1	Model A-2
B	$\bar{\pi}^{1\|0} = 10.98$	9.94 (7.92, 12.10)	9.96 (7.89, 12.38)	-
	$\bar{\pi}^{0\|1} = 8.69$	11.04 (8.27, 14.20)	10.04 (7.49, 12.95)	-
	$\bar{\pi}^{1\|0} = 20.19$	19.59 (17.03, 22.17)	17.37 (15.02, 19.65)	-
	$\bar{\pi}^{0\|1} = 21.05$	21.54 (17.97, 25.40)	17.54 (14.15, 21.02)	-

[a] In percentage points.

Table 7: Estimates of γ_1 and γ_2 under misspecificied linear predictors in the misreport model

Estimator	$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{2,0}$	$\gamma_{2,1}$	$\gamma_{2,2}$
True value[a]	-1.50	0.05	1.25	-0.20	0.05	-1.25
Specification 1	-1.45 (-1.63, -1.29)	-	1.20 (0.98, 1.42)	-0.12 (-0.34, 0.10)	-	-1.13 (-1.38, -0.85)
Specification 2	-1.44 (-1.64, -1.26)	-0.07 (-0.26, 0.11)	1.20 (1.00, 1.43)	-0.20 (-0.45, 0.08)	0.01 (-0.33, 0.35)	-1.11 (-1.40, -0.82)
Specification 3	-1.43 (-1.64, -1.26)	0.03 (-0.15, 0.20)	1.15 (0.89, 1.41)	-0.16 (-0.49, 0.15)	0.04 (-0.19, 0.28)	-1.20 (-1.63,-0.75)
Specification 4	-1.56 (-1.89, -1.21)	-0.02 (-0.15, 0.11)	1.19 (0.98, 1.42)	-0.21 (-0.68, 0.23)	0.08 (-0.09, 0.24)	-1.12 (-1.41, -0.85)

[a] $\bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.20$.

Table 8: Estimates of γ_1 and γ_2 under misspecificied error distributions in the misreport model

Error distribution	Estimator	$\gamma_{1,0}$	$\gamma_{1,1}$	$\gamma_{1,2}$	$\gamma_{2,0}$	$\gamma_{2,1}$	$\gamma_{2,2}$
Logistic[a]	True value	-1.75	0.65	1.30	-0.75	0.20	-1.30
	Propose method	-1.04 (-1.19, -1.88)	0.40 (0.27, 0.52)	0.77 (0.56, 1.00)	-0.43 (-0.72, -0.16)	0.17 (-0.01, 0.33)	-0.87 (-1.13, -0.59)
Bimodal[b]	True value	-1.60	0.50	1.30	-1.00	0.50	-1.30
	Propose method	-2.28 (-2.65, -1.94)	0.79 (0.60, 0.99)	1.99 (1.56, 2.43)	-1.27 (-1.70, -0.88)	0.73 (0.47, 1.04)	-2.11 (-2.74, -1.61)
Heteroskedastic[c]	True value	-2.05	0.95	0.10	-1.50	-2.50	-0.70
	Propose method	-0.95 (-1.10, -0.81)	0.82 (0.67, 0.98)	-0.10 (-0.35, 0.16)	-0.05 (-0.45, 0.35)	-3.00 (-3.73, -2.29)	-1.11 (-1.65, -0.55)

[a] $\bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.20$.
[b] $\bar{\pi}^{1|0}, \bar{\pi}^{0|1} \approx 0.10$.
[c] $\bar{\pi}^{1|0} \approx 0.20, \bar{\pi}^{0|1} \approx 0.10$.

B. DESCRIPTION OF THE DATASET USED FOR THE ANALYSIS OF TURNOUT MISREPORTING

Table 9: Vote misreporting in 1978–1990 ANES[a]

Election	$P(\tilde{y}_i = 1 \mid y_i = 0)$	$P(y_i = 0 \mid \tilde{y}_i = 1)$	$P(\tilde{y}_i = 0 \mid y_i = 1)$	$P(y_i = 1 \mid \tilde{y}_i = 0)$
1978	23.27	24.55	3.02	2.84
1980	24.48	16.52	0.58	1.37
1984	38.83	13.63	0.22	1.70
1986	31.55	17.70	0.66	1.40
1988	36.30	14.63	1.06	7.10
1990	26.83	16.83	3.67	6.46

[a] In percentage points.

B.1. *Variables used in the turnout model*

1. Indicators for demographic, socio-economic and political characteristics

   Age: 1 if $Age < 30$; 2 if $30 \leq Age < 45$; 3 if $45 \leq Age < 60$; 4 if $Age \geq 60$.

   Church Attendance: Frequency of church attendance. Coding: 1 if never; 2 if a few times a year; 3 if once or twice a month; 4 if every week or almost every week.

   Education: Highest grade of school or year of college completed. Coding: 1 if 8 grades or less; 2 if 9–12 grades with no diploma or equivalency; 3 if 12 grades, diploma or equivalency; 4 if some college; 5 if college degree.

   Female: 1 if the respondent is female, 0 if male.

   Home owner: 1 if the respondent owns his house, 0 otherwise.

   Income: Household income. Coding: 1 if 0–16th percentile; 2 if 17h–33d percentile; 3 if 34th–67th percentile; 4 if 68th–95th percentile; 5 if 96th–100th percentile.

   Non-white: 0 if white, 1 otherwise.

   Party Identification: -1 for Democrats, 0 for Independents, 1 for Republicans.

   Partisan Strength: Coded on a four-point scale ranging from 1 for pure independents to 4 for strong partisans.

2. Additional covariates to account for misreporting

   Alone: 1 if the respondent was interviewed alone, 0 otherwise.

   Uncooperative: Respondent's level of cooperation in the interview, as evaluated by the interviewer. Coding: 1 if very good; 2 if good; 3 if fair; 4 if poor; 5 if very poor.

   Sincerity: How sincere did the respondent seem to be in his/her answers, as evaluated by the interviewer. Coding: 1 if often seemed insincere; 2 if usually sincere; 3 if completely sincere.

In order to reduce the correlation between the parameters and to accelerate convergence and mixing of the Gibbs sampling algorithm, all variables where centered at their mean values (Gu 2006).

Table 10: Rates of nonresponse for the variables included in the voter turnout models

Variable	**1978-1990 validated ANES**	**1992 ANES**
Age	2.07	0.00
Church Attendance	13.20	33.72
Education	0.80	2.61
Female	4.28	0.00
Income	13.58	10.66
Non-white	4.41	1.41
Home owner	0.70	6.44
Partisan Strength	4.44	0.56
Party Identification	2.60	0.36
Alone	4.55	1.57
Cooperation	4.49	0.16
Sincerity	0.47	0.24
Reported turnout	6.12	9.30
Total sample	11,632	2,485
Complete-case sample	6,411	1,206

REFERENCES

Abelson, Robert P., Elizabeth F. Loftus and Anthony G. Greenwald. 1992. *Attempts to Improve the Accuracy of Self-Reports of Voting.* Russel Sage, New York: J. M. Tanur (ed.) Questions about Questions: Inquiries into the Cognitive Bases of Surveys, pp. 13853.

Abramson, Paul R. and William Claggett. 1984. "Race-Related Differences in Self-Reported and Validated Turnout." *Journal of Politics* 46:719–738.

Abramson, Paul R. and William Claggett. 1986*a*. "Race-Related Differences in Self-Reported and Validated Turnout in 1984." *Journal of Politics* 48:412–422.

Abramson, Paul R. and William Claggett. 1986*b*. "Race-Related Differences in Self-Reported and Validated Turnout in 1986." *Journal of Politics* 51:397–408.

Abramson, Paul R. and William Claggett. 1991. "Race-Related Differences in Self-Reported and Validated Turnout in the 1988 Presidential Election." *Journal of Politics* 53:186–197.

Abrevaya, Jason and Jerry A. Hausman. 1999. "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells." *Mimeo, Department of Economics, Massachussets Institute of Technology* .

Aigner, Dennis J. 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics* 1:49–60.

Aldrich, John H. 1993. "Rational Choice and Turnout." *American Journal of Political Science* 37(1):246–278.

Alvarez, R. Michael. 1997. *Information and Elections.* Michigan: The University of Michigan Press.

Anderson, Brady A. and Brian D. Silver. 1986. "Measurement and Mismeasurement of the Validity of the Self-Reported Vote." *American Journal of Political Science* 30:771–785.

Ansolabehere, Stephen and Eitan Hersh. 2008. *Vote Validation in the 2006 CCES.* Mimeo.

Battistin, Erich. 2003. *Errors in Survey Reports of Consumption Expenditures.* London: Working Paper 0307, Institute for Fiscal Studies.

Belli, Robert F., Michael W. Traugott, Margaret Young and Katherine A. McGonable. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63(1):90–108.

Belli, Robert F., Michael W. Traugott and Matthew N. Beckman. 2001. "What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies." *Journal of Official Statistics* 17(4):479–498.

Belli, Robert, Santa Traugott and Steven Rosenstone. 1994. "Reducing Over-Reporting of Voter Turnout: An Experiment Using a 'Source Monitoring' Framework." *ANES Technical Report Series* (010153).

Bernstein, Robert, Anita Chadha and Robert Montjoy. 2001. "Overreporting Voting. Why it Happens and Why it Matters." *Public Opinion Quarterly* 65(1):22–44.

Bollinger, Chris R. 1996. "Bounding Mean Regressions When a Binary Regressor is Mismeasured." *Journal of Econometrics* 73:387–399.

Bound, John, Charles Brown and Nancy Mathiowetz. 2001. *Measurement Error in Survey Data.* Elsevier, North Holland: J. Heckman and E. Leamer (eds.) Handbook of Econometrics, Vol. 5, pp. 3705-3843.

Bross, Irwin. 1954. "Misclassification in 2 X 2 Tables." *Biometrics* 10:478–486.

Burden, Barry C. 2000. "Voter Turnout and the National Election Studies." *Political Analysis* 8(4):389–398.

Cahalan, Don. 1968. "Correlates of Respondent Accuracy in the Denver Validity Survey." *Public Opinion Quarterly* 32(3):607–621.

Carroll, Raymond J., David Ruppert and Leonard A. Stefanski. 1995. *Measurement error in nonlinear models.* London: Chapman and Hall.

Casella, George and Edward L. George. 1992. "Explaining the Gibbs Sampler." *The American Statistician* 46(3):167–174.

Cassel, Carol A. 2003. "Overreporting and Electoral Participation Research." *American Politics Research* 31(1):81–92.

Chen, Ming-Hui, Joseph G. Ibrahim and Qi-Man Shao. 2000. "Power prior distributions for generalized linear models." *Journal of Statistical Planning and Inference* 84:121–137.

Chen, Qingxia and Joseph G Ibrahim. 2006. "Semiparametric Models for Missing Covariate and Response Data in Regression Models." *Biometrics* 62:177–184.

Chen, Qingxia, Joseph G. Ibrahim, Ming-Hui Chen and Pralay Senchaudhuri. 2008. "Theory and inference for regression models with missing responses and covariates." *Journal of Multivariate Analysis* 99:1302–1331.

Chen, T. Timothy. 1979. "Log-Linear Models for Categorical Data With Misclassification and Double Sampling." *Journal of the American Statistical Association* 74(366):481–488.

Chib, Siddharta and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings algorithm." *American Statistician* 49(4):327–335.

Christin, Thomas and Simon Hug. 2004. "Methodological Issues in Studies of Conflict Processes: Misclassifications and Endogenous Institutions." *Paper presented at the Annual Meeting of the American Political Science Association, Chicago, September 2-5.* .

Clausen, Aage. 1968. "Response Validity: Vote Report." *Public Opinion Quarterly* 32:588–606.

Davidov, Ori, David Faraggi and Benjamin Reiser. 2003. "Misclassification in Logistic Regression with Discrete Covariates." *Biometrical Journal* 5:541–553.

Drake, Christiana. 1993. "Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect." *Biometrics* 49:1231–1236.

Duan, Yuyan. 2005. *A Modified Bayesian Power Prior Approach with Applications in Water Quality Evaluation*. Blacksburg, VA: Doctoral Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University.

Dunson, David B. and Kenneth R. Tindall. 2000. "Bayesian Anaysis of Mutational Spectra." *Genetics* 156:1411–1418.

Dustmant, Christian and Arthur van Soest. 2004. "An analysis of speaking fluency of immigrants using ordered response models with classification errors." *Journal of Business and Economic Statistics* 22:312–321.

Fullerton, Andrew S., Jeffrey C. Dixon and Casey A. Borch. 2007. "Bringing Registration into Models of Vote." *Public Opinion Quarterly* 71(4):649–660.

Garthwaite, Paul H., Joseph B. Kadane and Anthony OHagan. 2004. *Elicitation*. Pittsburgh, PA: Technical Report 808, Department of Statistics, Carnegie Mellon University, Department of Statistics.

Gelfland, Alan E. and Adrian F. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(410):398–409.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference for iterative simulation using multiple sequences." *Statistical Science* 7:457–472.

Gelman, Andrew, Gary King and Chuanhai Liu. 1998. "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys." *Journal of the American Statistical Association* 93(443):846–857.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Boca Raton: Chapman and Hall.

Gems, Barbara, Dhirendra Ghaosh and Robert Hitlin. 1982. "A Recall Experiment: Impact of Time on Recall of Recreational Fishing Trips." *Proceedings of the Section on Survey Research Methods* 7:168–173.

Gilks, Walter R., Sylvia Richardson and David J. Spiegelhalter. 1996. *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Gu, Yuanyuan. 2006. *Misclassification of the Dependent Variable in Binary Choice Models*. Australia: Masters' Dissertation, School of Economics, University of the New South Wales.

Härdle, Wolfgang. 1990. *Applied nonparametric regression*. New York: Cambridge University Pres.

Haukka, Jari K. 1995. "Correction for covariate measurement error in generalized linear model - a bootstrap approach." *Biometrics* 51:1127–1132.

Hausman, Jerry A., Jason Abrevaya and Fiona M. Scott-Morton. 1998. "Misclassification of the dependent variable in a discrete response setting." *Journal of Econometrics* 87(2):239–269.

Hernanz, Virginia, Franck Malherbet and Michelle Pellizzari. 2004. "Take-up of Welfare Benefits in OECD Countries: A Review of the Evidence." *OECD Social, Employment and Migration Working Papers* (17).

Highton, Benjamin. 2004. "Self-reported versus Proxy-reported Voter Turnout in the Current Population Survey." *Public Opinion Quarterly* 69(1):113–123.

Hill, Kim Q. and Patricia A. Hurley. 1984. "Nonvoters in Voters' Clothing: The Impact of Voting Behavior Misreporting on Voting Behavior Research." *Social Science Quarterly* 65:199–206.

Horowitz, Joel L. 1993. *Semiparametric and Nonparametric Estimation of Quantal Response Models*. Elsevier, Amsterdam: G.S. Maddala, C. R. Rao, and H. D. Vinod (eds), Handbook of Statistics, Vol. 11, pp. 45–72.

Horowitz, Joel L. and Charles F. Manski. 1995. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63(2):281–302.

Horton, Nicholas J. and Ken P. Kleinman. 2007. "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models." *The American Statistician* 61(1):79–90.

Hu, Yingyao. 2008. "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics* 144:27–61.

Hu, Yingyao and Geert Riddert. 2007. *Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information*. Mimeo.

Huang, Lan, Ming-Hui Chen and Joseph G. Ibrahim. 1999. "Bayesian Analysis for Generalized Linear Models with Nonignorably Missing Covariates." *Biometrics* 61:767–780.

Ibrahim, Joseph G., Luise M. Ryan and Ming-Hui Chen. 1998. "Using Historical Controls to Adjust for Covariates in Trend Tests for Binary Data." *Journal of the American Statistical Association* 93(444):481–488.

Ibrahim, Joseph G. and Min-Hui Chen. 2000. "Power Prior Distributions for Regression Models." *Statistical Science* 15(1):46–60.

Ibrahim, Joseph G., Ming-Hui Chen and Stuart R. Lipsitz. 2002. "Bayesian Methods for Generalized Linear Models With Covariates Missing at Random." *Canadian Journal of Statistics* 30:55–78.

Ibrahim, Joseph G., Ming-Hui Chen, Stuart R. Lipsitz and Amy H. Herring. 2005. "Missing-Data Methods for Generalized Linear Models: A Comparative Review." *Journal of the American Statistical Association* 100(469):332–346.

Ibrahim, Joseph G. and Stuart R. Lipsitz. 1996. "Parameter Estimation from Incomplete Data in Binomial Regression When the Missing Data Mechanism is Nonignorable." *Applied statistics* 52(2):1071–1078.

Ibrahim, Joseph G., Stuart R. Lipsitz and Min-Hui Chen. 1999. "Missing Covariates in Generalized Linear Models when the Missing Data Mechanism Is Non-ignorable." *Journal of the Royal Statistical Society B* 61(1):173–190.

Ibrahim, Joseph G., Stuart R. Lipsitz and Nick Horton. 2001. "Using auxiliary data for parameter estimation with non-ignorably missing outcomes." *Applied statistics* 50(3):361–373.

Katosh, John P. and Michael W. Traugott. 1981. "The Consequences of Validated and Self-Reported Voting Measures." *Public Opinion Quarterly* 45:519–535.

Kuha, Jouni. 1994. "Corrections for exposure measurement error in logistic regression models with an application to nutritional data." *Statistics in Medicine* 13:1135–1148.

Leighley, Jan E. and Jonathan Nagler. 1984. "Individual and Systemic Influences on Turnout: Who Votes?" *Journal of Politics* 54:718–740.

Lewbel, Arthur. 2000. "Identification of the Binary Choice Model with Misclassification." *Econometric Theory* 16:603–609.

Lipsitz, S.R. and J.G. Ibrahim. 1996. "A Conditional Model for Incomplete Covariates in Parametric Regression Models." *Biometrika* 83:916–92.

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical analysis with missing data.* New York: Wiley.

Little, Roderick J. and Yongxiao Wang. 1996. "Pattern-Mixture Models for Multivariate Incomplete Data With Covariates." *Biometrics* 52:98–111.

Loftus, Elizabeth F. 1975. "Leading Questions and the Eyewitness Report." *Cognitive Psychology* 7:145–177.

Manski, Charles F. 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator." *Journal of Econometrics* 82:46–51.

Marshall, R. J. 1990. "Validation study methods for estimating proportions and odds ratios with misclassified data." *Journal of Clinical Epidemiology* 43.

Mcdonald, Michael P. 2007. "The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics." *Public Opinion Quarterly* 71(4):588–602.

McGlothlin, Anna, James D. Stamey and John W. Seaman. 2008. "Binary Regression with Misclassified Response and Covariate Subject to Measurement Error: a Bayesian Approach." *Biometrical Journal* 50:123–134.

McInturff, Pat, Wesley O. Johnon, David Cowling and Ian A. Gardner. 2004. "Modelling risk when binary outcomes are subject to error." *Statistics in Medicine* 23(7):1095–1109.

Miller, Mungo. 1952. "The Waukegan Study of Voter Turnout Prediction." *Public Opinion Quarterly* 32:588–606.

Molinari, Francesca. 2003. *Contaminated, Corrupted and Missing Data.* Evanston, IL: Doctoral Dissertation, Department of Economics, Northwestern University.

Morrissey, Mary J. and Donna Spiegelman. 1999. "Matrix methods for estimating odds ratios with misclassified exposure data: Extensions and comparisons." *Biometrics* 55(398):338–344.

Murphy, Kevin M. and Richard H. Topel. 1985. "Estimation and Inference in Two Step Econometric Models." *Journal of Business and Economic Statistics* 3:370–379.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38(1):230–255.

Neuhaus, John M. 1999. "Bias and efficiency loss due to misclassified responses in binary regression." *Biometrika* 86(4):843–855.

Parry, Hugh and Helen Crossley. 1950. "Validity of Responses to Survey Questions." *Public Opinion Quarterly* 14:61–80.

Paulino, Carlos D., Paulo Soares and John Neuhaus. 2003. "Binomial Regression with Misclassification." *Biometrics* 59:670–675.

Plummer, Martyn. 2009. *JAGS version 1.03 manual.* www-ice.iarc.fr/˜/software/jags/.

Poterba, James M. and Lawrence H. Summers. 1986. "Reporting Errors and Labor Market Dynamics." *Econometrica* 54:1319–1338.

Poterba, James M. and Lawrence H. Summers. 1995. "Unemployment benefits and labor market transitions: a multinomial logit model with errors in classification." *The Review of Economics and Statistics* 77(2):207–216.

Prescott, Gordon J. and Paul H. Garthwaite. 2002. "Simple Bayesian Analysis of Misclassified Binary Data with a Validation Substudy." *Biometrics* 58:454–458.

Prescott, Gordon J. and Paul H. Garthwaite. 2005. "Bayesian analysis of misclassified binary data from a matched case-control study with a validation sub-study." *Statistics in Medicine* 24(3).

Reilly, Marie and Margaret S. Pepe. 1995. "A mean score method for missing and auxiliary covariate data in regression models." *Biometrika* 82:299–314.

Robert, Christian P. and George Casella. 2004. *Monte Carlo Statistical Methods.* New York: Springer.

Robins, James M. and Andrea G. Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data." *Journal of the American Statistical Association* 90:122–129.

Robins, James M., Andrea G. Rotnitzky and Lue P. Zhao. 1994. "Estimation of regression coefficients when some regressors are not always observed." *Journal of the American Statistical Association* 89:846–866.

Rotnitzky, Andrea G. and James M. Robins. 1995. "Semiparametric Regression Estimation in the Presence of Dependent Censoring." *Biometrika* 82:805–820.

Rubin, Donald B. 1976. "Inference and missing data." *Biometrika* 63:581–592.

Schafer, Joseph L. and John W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7(2):147–177.

Sigelman, Lee. 1982. "The Nonvoting Voter in Voting Research." *American Journal of Political Science* 26:47–56.

Silver, Brian D., Barbara A. Anderson and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80:613–624.

Spiegelhalter, David J., Andrew Thomas and Nicky G. Best. 2003. *WinBUGS, Version 1.4. User Manual.* Cambridge, UK: Medical Research Council Biostatistics Unit, University of Cambridge.

Tanner, Martin A. and Wing Hung Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82(398):528–540.

Thürigen, Dorothee, Donna Spiegelman, Maria Blettner, Cartsten Heuer and Hermann Brenner. 2000. "Measurement error correction using validation data: a review of methods and their applicability in case-control studies." *Statistical Methods in Medical Research* 9:447–474.

Viana, M. A. 1994. "Bayesian small-sample estimation of misclassified multinomial data." *Biometrics* 50(1):237–243.

Weir, Blair T. 1975. "The Distortion of Voter Recall." *American Journal of Political Science* 19:53–62.

Wiegmann, Douglas A. 2005. *Developing a Methodology for Eliciting Subjective Probability Estimates During Expert Evaluations of Safety Interventions: Application for Bayesian Belief Networks.* Hampton, VA: Final Technical Report AHFD-05-13/NASA-05-4, NASA Langley Research Center.

Wolfinger, Raymond E. and Steven J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.

Zhao, Zong. 2008. "Sensitivity of Propensity Score Methods to the Specifications." *Economics Letters* 98(3):309–319.