

Correcting for Survey Nonresponse Using Variable Response Propensity

Michael PERESS

All surveys with less than full response potentially suffer from nonresponse bias. Poststratification weights can only correct for selection into the sample based on observables whose distribution is known in the population. Variables such as gender, race, income, and region satisfy this requirement because they are available from the U.S. Census Bureau, but poststratification based on these variables may not eliminate nonresponse bias. I develop an approach for correcting for nonignorable nonresponse bias. Survey respondents can be classified by their “response propensity.” Proxies for response propensity include the number of attempted phone calls, indicators of temporary refusal, and interviewer-coded measures of cooperativeness. We can then learn about the population of nonrespondents by extrapolating from the low-propensity respondents. I apply this new estimator to correct for unit nonresponse bias in the American National Election Study and in a CBS/New York Times preelection poll. I find that nonresponse bias can be a serious problem, particularly for items that relate to political participation. I find that my method is successful in substantially reducing nonresponse bias.

KEY WORDS: Continuum of resistance; Nonresponse bias; Survey; Weighting.

1. INTRODUCTION

In an ideal world, a survey would consist of a random sample of the population. Each individual in the population of interest would have an equal chance of being interviewed, or at minimum, this probability would be known. If this were the case, simple averages could be used to estimate most quantities of interest. But these assumptions are almost always violated.

Survey methodologists have devised fairly effective methods for generating a random sample of households. Random digit dialing can produce a representative sample of phone lines. By adjusting for the number of phone lines in a household, we can obtain a representative sample of households. By adjusting for the size of the household, we can obtain a representative sample of adults. The missing link is that not every selected adult will be reachable and willing to participate.

Along with less than complete response comes the possibility of nonresponse bias—the responding portion of the population differs from the nonresponding portion. Under these conditions, simple averages produce biased estimates of the population parameters of interest. Weighting the data such that the distribution of observables in the sample matches the distribution in the population is often used to correct for nonresponse bias. However, this technique will produce biased estimates as well if the data are not “missing at random” [in the sense of Rubin (1987)]. I develop an approach that directly corrects for nonignorable unit nonresponse bias. My approach classifies survey respondents by their “response propensity,” and extrapolates from the low-propensity respondents to the nonrespondents.

I apply the “variable response propensity estimator” to correct for nonresponse bias in the American National Election Study (ANES). This dataset allows evaluation of the estimator in an environment where the “truth” is known, from election results and the voter validation studies. The ANES also provides

multiple useful measures of response propensity—the number of calls or visits to a household, whether there was an initial refusal, interviewer-coded measures of cooperativeness, etc.

In the years I study, the ANES overestimates turnout by 9–12%, even after measurement error is corrected for. When poststratification weights are applied to correct for nonresponse bias, the discrepancy drops to between 6% and 9%. The results indicate that my approach is quite successful in reducing nonresponse bias. My estimator almost always provides a better estimate than a demographic-weighted proportion. Some of the measures of response propensity almost completely eliminate nonresponse bias. The number of calls to a household is the least effective measure. Interviewer-coded measures of cooperativeness and interest in the interview are the most successful.

I apply the new estimator to diagnose and correct for nonresponse bias in a number of other items in the ANES and for a number of items in a CBS/New York Times preelection poll. Items that relate to ideology generally exhibit small to moderate amounts of nonresponse bias. Items that relate to political participation exhibit moderate to large amounts of nonresponse bias. In particular, we will overestimate political participation because political surveys such as the ANES and preelection polls naturally over-sample individuals who are interested in politics. Unlike demographic weighting, my method is capable of correcting for nonignorable nonresponse bias, which is present in many of the survey items social scientists deal with.

2. NONRESPONSE BIAS

Survey researchers have become increasingly worried about unit nonresponse bias, in a large part due to decreasing response rates. Efforts to compensate for a decreasing willingness of households to participate include making multiple calls, attempting to convert initial refusers, and providing cash incentives to respondents or interviewers (Curtin, Presser, and Singer 2000; Brick et al. 2005). Increased effort can indeed lead to increased response rates (Keeter et al. 2000; Brick et al. 2003), yet

Michael Peress is Assistant Professor of Political Science, Department of Political Science, University of Rochester, Rochester, NY 14627 (E-mail: mperess@mail.rochester.edu). This article benefited from comments and suggestions from Roeland Beerten, Robert Erikson, Garrett Glasgow, Dick Niemi, Lynda Powell, Curt Signorino, three anonymous referees, the associate editor, and seminar participants at the University of Rochester, the American Association for Public Opinion Research conference (Montreal, 2006), and the American Political Science Association conference (Chicago, 2007).

the evidence is quite mixed as to whether marginal increases in response rates will actually reduce nonresponse bias.

Keeter et al. (2000) found that most variables yielded similar estimates in their rigorous and nonrigorous surveys, despite dramatically different response rates. Curtin, Presser, and Singer (2000) found that excluding late responders from the analysis did not lead to many differences in estimates of the Index of Consumer Sentiment. Montaquila et al. (2008) found few differences when no refusal conversions were attempted and fewer callbacks were made. Groves and Peytcheva (2008) performed a meta-analysis of 59 surveys, and determined that there is a very weak relationship between the nonresponse rate and nonresponse bias.

However, a number of studies found that significant differences existed between early and late responders (Ellis, Endo, and Armer 1970; Dunkelberg and Day 1973; Traugott 1987; Teitler, Reichman, and Sprachman 2003; Voigt, Koepsell, and Daling 2003), suggesting that in these cases, stopping the surveys earlier would have led to increased nonresponse bias. Moreover, a number of studies found significant differences between amicable respondents and initial refusers (Hawkins 1975; O'Neil 1979; Stinchcombe, Jones, and Sheatsley 1981; Fitzgerald and Fuller 1982; Smith 1984; Voigt, Koepsell, and Daling 2003), indicating that decreased effort would have led to different results. Collectively, these results suggest that marginal increases in response rates may not appreciably reduce nonresponse bias, and nonresponse bias may still be present even in the most rigorously conducted survey. Direct methods for diagnosing and correcting for nonresponse bias are essential.

The most common approach to correcting for unit nonresponse involves applying poststratification weights (Groves et al. 2002, 2004). In order to implement this correction method, two conditions must be met. First, we must be able to identify and measure those variables that determine selection into the group of respondents. Second, the distribution of those variables in the population of interest must be known. For example, it is well known that the elderly are more likely to participate in surveys (Brehm 1993). To correct for this fact, one could weight the young more heavily when computing sample averages.

Post-stratification weights are extremely useful and are widely employed (Voss, Gelman, and King 1995), but there are some limitations. This approach will not eliminate nonresponse bias if the data are not missing at random (or put differently, if there is selection on unobservables). A survey may over-represent households that are interested in the topic of the survey (Groves, Presser, and Dipko 2004). A political survey, in particular, may over-represent individuals who are interested in politics, and consequently may overestimate the proportion of American adults who vote, attend a campaign event, read the newspaper, etc. One could not easily correct this problem using poststratification weights since the proportion of adults who are interested in politics would not be known (it could be estimated using a separate survey, but this survey may itself be subject to the same type of nonresponse bias).

2.1 Models of Survey Nonresponse

Insight into understanding and correcting for unit nonresponse bias can be obtained from the continuum of resistance

model and the classes model. The continuum of resistance model (Ellis, Endo, and Armer 1970; Dunkelberg and Day 1973; Filion 1975, 1976; Lin and Schaeffer 1995; Teitler, Reichman, and Sprachman 2003; Biemer and Link 2008) posits that voters differ in their response propensity. Those in the population with low-response propensity are less likely to respond. We can infer the variables of interest for the nonrespondents by extrapolating from the low-propensity respondents. The "classes" model (O'Neil 1979; Stinchcombe, Jones, and Sheatsley 1981; Smith 1984; Lin and Schaeffer 1995; Curtin, Presser, and Singer 2000; Montaquila et al. 2008) instead posits that there are groups of respondents who resemble the nonrespondents. For example, refusers may resemble temporary refusers and unlocated individuals may resemble hard-to-locate individuals.

Both models require measuring response propensity, and response propensity may be measured in a number of ways. The most common measure is the number of contacts with the selected individual. These contacts may take the form of callbacks (Hawkins 1975; Potthoff, Manton, and Woodbury 1993; Voigt, Koepsell, and Daling 2003), home visits (Ellis, Endo, and Armer 1970; Filion 1976), or mailings (Filion 1975; Drew and Fuller 1980, 1981). An alternative measure is whether the respondent was an initial refuser. An initial refuser may be "converted" through additional calls, persuasion letters, or monetary incentives.

Alternatively, one could consider interviewer-coded measures of the respondent's cooperativeness and interest in the interview. Though these measures are available less readily, they may be useful as well. Self reports of willingness to participate in future surveys and measures based on the degree of item nonresponse provide other possible measures.

2.2 Correction Methods

Both the classes and continuum of resistance models have led to methods for correcting for nonignorable nonresponse bias. O'Neil (1979), Fitzgerald and Fuller (1982), and Lin and Schaeffer (1995) apply the classes method to correct for nonresponse bias. Of these, only Lin and Schaeffer compare their results to a known benchmark.

Filion (1975, 1976), Drew and Fuller (1980, 1981), Potthoff, Manton, and Woodbury (1993), and Biemer and Link (2008) apply correction methods based on the continuum of resistance model. Filion (1975, 1976) propose fitting a regression line where the dependent variable is the survey measure and the independent variable is the wave number. The value of the dependent variable for the nonrespondents becomes the predicted value if one more wave had been run. While Filion (1975) did not have a benchmark against which to evaluate the estimates, Filion (1976) found that this method produced appropriate estimates (though the original estimates also exhibited little nonresponse bias due to the small number of nonrespondents).

Drew and Fuller (1980, 1981) propose a method for extrapolating from the respondents to the nonrespondents based on the number of callbacks. The procedure accounts for the fact that some groups (e.g., old female respondents) respond earlier than other groups of respondents. Their method uses this relationship to extrapolate to the group makeup of the nonrespondents. Biemer and Link (2008) extend Drew and Fuller's method to

differentiate between noncontacts and refusals. [Potthoff, Manton, and Woodbury \(1993\)](#) present an alternative method for extrapolating to nonrespondents based on distribution of the number of callbacks.

[Brehm \(1993, 1999\)](#) considers an alternative method for correcting for nonresponse bias in regression coefficients, using a [Heckman \(1979\)](#) sample selection model. One of his models employs survey administration variables, including the number of callbacks and whether the respondent initially refused.

My method follows the spirit of the continuum of resistance models, but differs in a number of important ways. First, I consider a number of different measures of response propensity, including measures not previously considered. The existing correction methods based on the continuum of resistance model have all focused on callbacks. I find that while corrections based on the number of callbacks and successful conversions provide an improvement over standard estimators, interviewer-coded measures of cooperativeness and interest in the interview are the most effective.

Second, my method incorporates observable covariates in the framework, retaining the benefits of poststratification while incorporating a correction for nonignorable nonresponse bias. The methods of [Filion \(1975, 1976\)](#) and [Potthoff, Manton, and Woodbury \(1993\)](#) do not incorporate such a correction (though it is possible that their methods could be adapted to incorporate observable covariates). The methods of [Drew and Fuller \(1980, 1981\)](#) and [Biemer and Link \(2008\)](#) allow the distribution of groups in the population to be extrapolated from the sample and are useful in situations in which demographic targets are not known.

Third, I consider a test of my method in a situation in which the “truth” is known. Few existing studies applying correction methods incorporate an instance where the truth is known, with the exceptions being [Ellis, Endo, and Armer \(1970\)](#), [Filion \(1976\)](#), [Lin and Schaeffer \(1995\)](#), and [Teitler, Reichman, and Sprachman \(2003\)](#).

Finally, unlike any of the existing extrapolation methods, my method applies well when response propensity is not measured on an interval scale. It is this fact that allows me to consider measures other than the number of callbacks. Moreover, even for this measure, it may be inappropriate to assume that nonresponse takes the value of $R + 1$ on this interval scale (where R is the maximum number of callbacks). Instead, my approach employs an ordered categorical response equation.

3. ESTIMATION OF THE POPULATION PROPORTION

In this section, I derive the variable response propensity estimator (VRPE), which can potentially correct for nonignorable nonresponse bias. My correction method will require specifying a joint statistical model of the outcome variable and response propensity, which I will now describe.

3.1 Model

Suppose that y_n is binary outcome variable. The outcome equation is given by $y_n^* = \alpha' \mathbf{x}_n + \varepsilon_n$ where $y_n = 1$ if $y_n^* < 0$ and $y_n = 0$ otherwise. Here, \mathbf{x}_n is a vector of regressors, ε_n is a mean zero disturbance term, and α is a vector of unknown parameters characterizing the outcome equation.

I let r_n denote the response category, and I assume that $r_n \in \{1, 2, \dots, R, R + 1\}$. Categories 1 through R denote different levels of measured response propensity with 1 indicating high response propensity and R indicating low response propensity. I let $r_n = R + 1$ denote nonresponse by individual n . The response equation is given by $r_n^* = \beta' \mathbf{z}_n + \eta_n$. Here, \mathbf{z}_n is a vector of regressors, η_n is a mean zero disturbance term, and β is a vector of unknown parameters characterizing the selection equation. I model r_n as an ordered probit equation. I assume that we observe $r_n = 1$ if $r_n^* < \theta_1$, $r_n = 2$ if $\theta_1 < r_n^* < \theta_2$, \dots , $r_n = R$ if $\theta_{R-1} < r_n^* < \theta_R$, and $r_n = R + 1$ if $r_n^* > \theta_R$. Here, θ_r denote the cutoffs that are used to classify respondents into response categories.

The framework assumes that we can differentiate among respondents in their response propensity. Measures that can be used for r_n include the number of callbacks, whether the respondent initially refused, and interviewer-coded cooperativeness. I assume that $R \geq 2$, or that there are at least two categories of respondents. For example, we may have interviewer-coded cooperativeness, with $r_n = 1$ indicating a cooperative respondent, $r_n = 2$ indicating an uncooperative respondent, and $r_n = 3$ indicating a nonrespondent.

I assume that $(y_n, r_n, \mathbf{x}_n, \mathbf{z}_n)$ are observed if $r_n \leq R$. Otherwise, the analyst observes only r_n . I let N denote the number of observations for which $r_n \leq R$ and let N_{miss} denote the number of observations for which $r_n = R + 1$. I assume that the observations are ordered such that $r_n \leq R$ for the first N observations and $r_n = R + 1$ for the last N_{miss} observations. I assume for simplicity that each unit in the population has an equal probability of entering the selected sample of $N + N_{miss}$ individuals. Individuals differ in their probability of entering the achieved sample of N respondents.

In addition to the sample, I assume that the analyst knows the distributions of \mathbf{x}_n and \mathbf{z}_n in the population. As such, these should be viewed as demographic characteristics whose distribution would be known from census data. The distributions of \mathbf{x}_n and \mathbf{z}_n are thus analogous to weighting targets widely employed by public and academic pollsters.

For simplicity, I will also assume that \mathbf{x}_n and \mathbf{z}_n are discrete random variables. Suppose that $\mathbf{x}_n = \tilde{\mathbf{x}}_j$ and $\mathbf{z}_n = \tilde{\mathbf{z}}_k$ with probability $p_{j,k}$, for $j = 1, \dots, J$ and $k = 1, \dots, K$. I let $p_j^x = \sum_{k=1}^K p_{j,k}$ and $p_k^z = \sum_{j=1}^J p_{j,k}$ denote the marginal distributions of \mathbf{x}_n and \mathbf{z}_n . I assume that ε_n and η_n are jointly normally distributed with mean 0 and variance 1, and have correlation ρ . Notice that the framework allows for both selection on observables (if \mathbf{x}_n and \mathbf{z}_n are not independent) and selection on unobservables (if $\rho \neq 0$).

The quantity of interest is $\pi = \Pr(y = 1)$ (which could be the president’s approval rating or the proportion of individuals who attended a campaign event). This quantity is equal to

$$\pi = \Pr(y = 1) = \sum_{j=1}^J \sum_{k=1}^K p_{j,k} \Pr(\alpha' \tilde{\mathbf{x}}_j + \varepsilon \leq 0)$$

$$= \sum_{j=1}^J \sum_{k=1}^K p_{j,k} \Phi(-\alpha' \tilde{\mathbf{x}}_j) = \sum_{j=1}^J p_j^x \Phi(-\alpha' \tilde{\mathbf{x}}_j).$$

The goal then is to obtain a consistent estimator of π , the population proportion.

3.2 Simple Estimators

The simplest estimator we could consider is the sample proportion (SP), $\hat{\pi}_1 = \frac{1}{N} \sum_{n=1}^N y_n$. This estimator will be a consistent estimator of π only under very strong assumptions. We would require that \mathbf{x}_n and \mathbf{z}_n are independent and that ε_n and η_n are independent. In other words, the data would have to be missing completely at random.

An alternative to the sample proportion is the weighted sample proportion (WSP). Define, $\hat{v}_j = \sum_{n=1}^N 1\{\mathbf{x}_n = \tilde{\mathbf{x}}_j\} y_n / \sum_{n=1}^N 1\{\mathbf{x}_n = \tilde{\mathbf{x}}_j\}$. Then the weighted sample proportion is given by $\hat{\pi}_2 = \sum_{j=1}^J p_j^x \hat{v}_j$. The weighted sample proportion is not in general consistent, but suppose, however, that ε_n and η_n are uncorrelated. In this case, the estimator will converge to the true parameter of interest if the unobservables are uncorrelated. This is equivalent to saying that the data are missing at random, or that we can fully control for selection using observables (Rosenbaum and Rubin 1983; Little 1986; Bethlehem 1988).

3.3 Variable Response Propensity Estimator

I present an estimator that is consistent even when ε_n and η_n are correlated. I will proceed by estimating the underlying model for $(y_n, r_n, \mathbf{x}_n, \mathbf{z}_n)$ using maximum likelihood. Computing the likelihood function requires characterizing the following probabilities:

$$\begin{aligned} \Pr(y_n = 0, r_n = r, \mathbf{x}_n = \tilde{\mathbf{x}}_j, \mathbf{z}_n = \tilde{\mathbf{z}}_k) \\ &= p_{j,k} \int_{\alpha' \tilde{\mathbf{x}}_j + \varepsilon \geq 0, \theta_r \leq \beta' \tilde{\mathbf{z}}_k + \eta \leq \theta_{r+1}} \phi(\varepsilon, \eta) d\varepsilon d\eta, \\ \Pr(y_n = 1, r_n = r, \mathbf{x}_n = \tilde{\mathbf{x}}_j, \mathbf{z}_n = \tilde{\mathbf{z}}_k) \\ &= p_{j,k} \int_{\alpha' \tilde{\mathbf{x}}_j + \varepsilon \leq 0, \theta_r \leq \beta' \tilde{\mathbf{z}}_k + \eta \leq \theta_{r+1}} \phi(\varepsilon, \eta) d\varepsilon d\eta, \\ \Pr(r_n = R + 1) &= \sum_{k=1}^K p_k^z \int_{\beta' \tilde{\mathbf{z}}_k + \eta \geq \theta_R} \phi(\varepsilon, \eta) d\varepsilon d\eta. \end{aligned}$$

Combining these, we can write the log-likelihood as

$$\begin{aligned} l(\xi) &= \sum_{n=1}^N \sum_{r=1}^R 1\{r_n = r, y_n = 0\} \\ &\quad \times \log \int_{\alpha' \mathbf{x}_n + \varepsilon \geq 0, \theta_r \leq \beta' \mathbf{z}_n + \eta \leq \theta_{r+1}} \phi(\varepsilon, \eta) d\varepsilon d\eta \\ &\quad + \sum_{n=1}^N \sum_{r=1}^R 1\{r_n = r, y_n = 1\} \\ &\quad \times \log \int_{\alpha' \mathbf{x}_n + \varepsilon < 0, \theta_r \leq \beta' \mathbf{z}_n + \eta \leq \theta_{r+1}} \phi(\varepsilon, \eta) d\varepsilon d\eta \\ &\quad + N_{miss} \log \sum_{k=1}^K p_k^z \int_{\beta' \tilde{\mathbf{z}}_k + \eta \geq \theta_R} \phi(\eta) d\eta, \end{aligned}$$

where $\xi = (\alpha, \beta, \rho, \theta)$ denotes the model parameters. I normalize $\theta_R = 0$ for identification purposes.

The likelihood function above does not admit a closed-form expression. In particular, evaluating the likelihood involves computing rectangles of the normal distribution. I compute

these integrals using the GHK method, which computes these integrals using simulation methods (Geweke, Keane, and Runkle 1994).

Recall that $\pi = \sum_{j=1}^J p_j^x \Phi(-\alpha' \tilde{\mathbf{x}}_j)$. This suggests that we estimate π using $\hat{\pi}_3 = \sum_{j=1}^J p_j^x \Phi(-\hat{\alpha}' \tilde{\mathbf{x}}_j)$. The variable response propensity estimator (VRPE) will be consistent, even in the presence of nonignorable nonresponse bias, provided that the underlying modeling framework is correct.

We can obtain a standard error estimate for $\hat{\pi}_3$ using the delta method (Green 2000). Notice that $\frac{\partial \hat{\pi}_3}{\partial \alpha_k} = -\sum_{j=1}^J p_j^x \times \tilde{x}_{j,k} \phi(-\hat{\alpha}' \tilde{\mathbf{x}}_j)$ so that the standard error is given by

$$se(\hat{\pi}_3) = \sqrt{\frac{1}{N + N_{miss}} \sum_{j_1=1}^J \sum_{j_2=1}^J p_{j_1}^x p_{j_2}^x \phi(-\hat{\alpha}' \tilde{\mathbf{x}}_{j_1}) \phi(-\hat{\alpha}' \tilde{\mathbf{x}}_{j_2}) \tilde{\mathbf{x}}_{j_1}' \hat{\mathbf{V}}_{\alpha} \tilde{\mathbf{x}}_{j_2}},$$

where $\hat{\mathbf{V}}_{\alpha}$ is the usual maximum likelihood estimator for the asymptotic variance of $\hat{\alpha}$.

It is important to note the assumptions that underlie this estimation procedure. The most important substantive assumption is that we can extrapolate based on response propensity. At first, this may seem like a rather strong assumption. Putting this in context, however, one should remember that *any* inference involving an incomplete sample involves extrapolation. When one applies a simple or weighted proportion, one is extrapolating from *all* respondents to the nonrespondents. Employing the classes method involves extrapolating from the low-propensity respondents to the nonrespondents. My method extrapolates based on the relationship between response propensity and the outcome variable.

I illustrate this in Figure 1. In this figure, I plot the outcome variable (e.g., voter turnout) by response propensity. The first

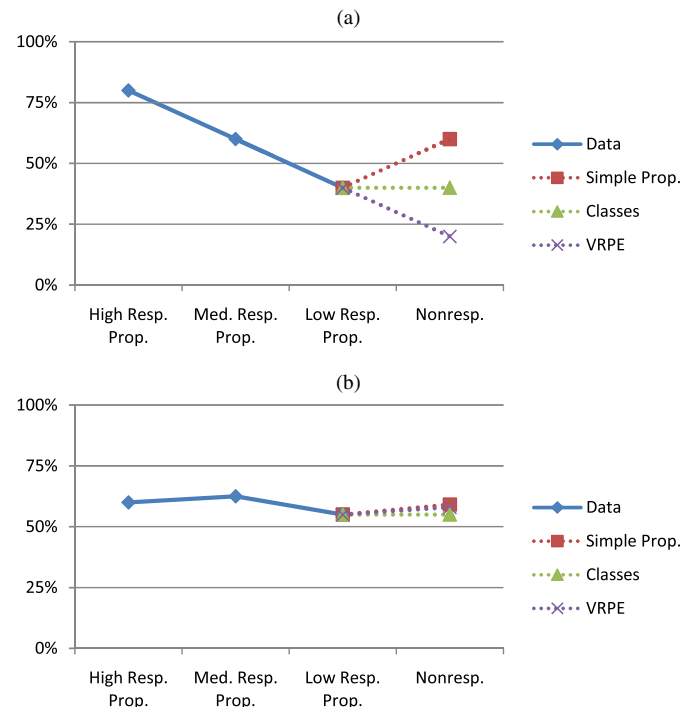


Figure 1. Extrapolating to the nonrespondents. (a) Strong relationship. (b) Weak relationship. The online version of this figure is in color.

panel indicates a strong negative relationship between the outcome variable and response propensity. As we can see, applying a simple proportion involves a rather peculiar form of extrapolation, requiring us to believe that the relationship between response propensity and the outcome variable experiences a sudden kink. The classes method provides a somewhat more sensible extrapolation, but the extrapolation employed by the variable response propensity estimator is the most sensible. In the event of a weak relationship, all three methods will lead to a similar estimate, and can be seen in the second panel of Figure 1. Consequently, my method does not introduce a new assumption, but replaces an existing assumption with a more reasonable one. Furthermore, in the event that the procedure finds no evidence of nonresponse bias (e.g., there is no linear pattern in the plot of the outcome variable against response propensity), the method will produce an estimate that will be nearly identical to the weighted sample proportion.

3.4 Identification

The model estimated here shares some similarities with the Tobit model and the sample selection model. Identification is known to be “tricky” in both cases. Both models are parametrically identified. Chamberlain (1986) demonstrates that a semiparametric selection model is formally identified, even in the absence of an exclusion restriction. At the same time, these types of models often preform poorly under misspecification (Arabmazar and Schmidt 1982; Goldberger 1983). Chamberlain (1986) suggests that his own result is not a convincing determination of identification due to his reliance on an “identification at infinity” argument, and suggests that further study is needed.

Collectively, these results suggests that formal results become quite tricky and are unlikely to resolve the underlying concerns. There is good reason to be more optimistic about the performance of the VRPE however. First, we observe more that one category of response among the respondents. This allows us to observe the relationship between the dependent variable and the response equation (something which is not possible for the sample selection model). Second, the parameters α , β , and ρ are nuisance parameters in our framework—we are ultimately interested in a function of α , which may be easier to estimate. I verify these institutions in the next section, where I employ Monte Carlo simulations to demonstrate that the VRPE is not very sensitive to alternative assumptions about the distributions of the disturbance terms or the form of the dependence between the disturbance terms.

4. MONTE CARLO RESULTS

In this section, I provide a monte carlo study of the variable response propensity estimator. The VRPE makes parametric assumptions about the distribution of the error terms in the model and the form of the dependence between the outcome and selection error terms—in particular, the errors are assumed to be drawn from a bivariate normal distribution. We would like to know that the VRPE is not substantially biased when reasonable departures from multivariate normal errors are considered.

To consider a “worst case scenario,” I assume that $\mathbf{z} = \mathbf{x}$. In this case, the ability to separate the determinants of the outcome variable from the determinants of response comes from

observing multiple response categories among the respondents. I assume that the variable \mathbf{x} takes on the values 0, 1, and 2, with probabilities 0.3, 0.3, and 0.4, respectively. I consider three choices for ρ . The first choice ($\rho = 0.0$) corresponds to the missing completely at random case. The second choice ($\rho = 0.2$) corresponds to a moderate amount of selection on unobservables, of the magnitude I find in the empirical applications reported later in the article. The third choice ($\rho = 0.5$) corresponds to a very large degree of selection on observables corresponding to substantial nonresponse bias for the SP and the WSP.

I chose the values $\alpha = [0, 1, 0.5]$ and $\beta = [0, 0.5, 0]$. I let $R = 3$ and $\theta = [-1, -0.5, 0]$ and I selected $N_{tot.} = N + N_{miss.} = 3000$. These values are, to a degree, arbitrary, but were chosen to replicate the typical environment a survey researcher would face. The parameter values correspond to a nonresponse rate of about 55% with a sample size of about 1300. The respondents are approximately equally divided between the three response categories.

In the first set of simulations, I considered a correctly specified model:

- Case 1: The errors are jointly normally distributed with variance 1 and correlation ρ .

To examine how the estimates respond to misspecification, I considered altering the marginal distributions of the errors terms and the form of the dependence between the errors terms. The marginal distributions can be altered while leaving the dependence structure intact by employing the Gaussian copula. I considered the following three cases:

- Case 2: The errors are generated according to a Gaussian copula with parameter ρ and the errors have logistic marginal distributions with variance 1.
- Case 3: The errors are generated according to a Gaussian copula and the errors terms have scaled chi-squared marginal distributions with 3 degrees of freedom, mean 0, and variance 1.
- Case 4: The errors are generated according to a Gaussian copula, the outcome error is logistic, and the selection error is chi-squared with 3 degrees of freedom.

Case 2 corresponds to a mild degree of misspecification—the logistic distribution looks quite similar to the normal. Case 3 considers error terms with a skewed distribution as thus corresponds to substantial misspecification. Case 4 considers what happens when the outcome and selection error terms have different distributions.

In the final two cases, I considered an alternative dependence structure:

- Case 5: The errors are generated according to a t copula with 5 degrees of freedom and dependence parameter ρ , the outcome error is logistic, and the selection error is chi-squared with 3 degrees of freedom.
- Case 6: The errors are generated according to a Clayton copula, the outcome error is logistic, and the selection error is chi-squared with 3 degrees of freedom.

Case 5 allows for stronger dependence in the tails than is assumed by the Gaussian copula, but preserves symmetry in

the dependence structure. Case 6 presents an extreme departure for bivariate normality because the Clayton copula exhibits dependence in the lower tail, but independence in the upper tail.

When we vary the speciation, we cannot hold the response rate constant without varying the other parameters in the model. While case 1 and case 2 have similar nonresponse rates (around 55%), case 3, case 4, case 5, and case 6 have nonresponse rates of around 45%. This difference occurs because the error term in the selection equation is right skewed. There is no closed form solution for the correlation in the Clayton copula, so in case 6, I set the dependence parameter so that the correlation is approximately 0.2 and 0.5.

The results are reported in Figure 2. All calculations are performed using $S = 100$ Monte Carlo replications. I employ the full sample proportion (FSP) as a benchmark for evaluating the estimators. The FSP averages y_n over both respondents and nonrespondents. This estimator is infeasible in any real application because we don't observe y_n for nonrespondents, but is a useful benchmark in this simulation study. Figure 2 reports the

distribution of the SP minus the FSP (in red or dark grey), the WSP minus the FSP (in blue or medium grey), and the VRPE minus the FSP (in green or light grey).

Consider first case 1. When $\rho = 0$, the data are missing at random, but not missing completely at random, in which case the SP should be biased and inconsistent and the WSP should be consistent. We see that the distribution of the SP is not centered at zero while the distribution of the WSP is centered at zero. The VRPE is centered at zero, but has a larger variance, reflecting the fact that if the data are missing at random, the VRPE is consistent, but inefficient relative the WSP.

As ρ is increased to 0.2 and 0.5, the WSP exhibits substantial bias because the data are not missing at random. The VRPE performs better—the bias is much smaller than the SP and the WSP. The WSP generally provides some small improvement over the SP. The VRPE exhibits some small amount of finite sample bias (which is statistically significantly different from zero in the $\rho = 0.5$ case) but the bias is small relative to the variance of the estimates and the VRPE provides a substantial improvement over the SP and the WSP.

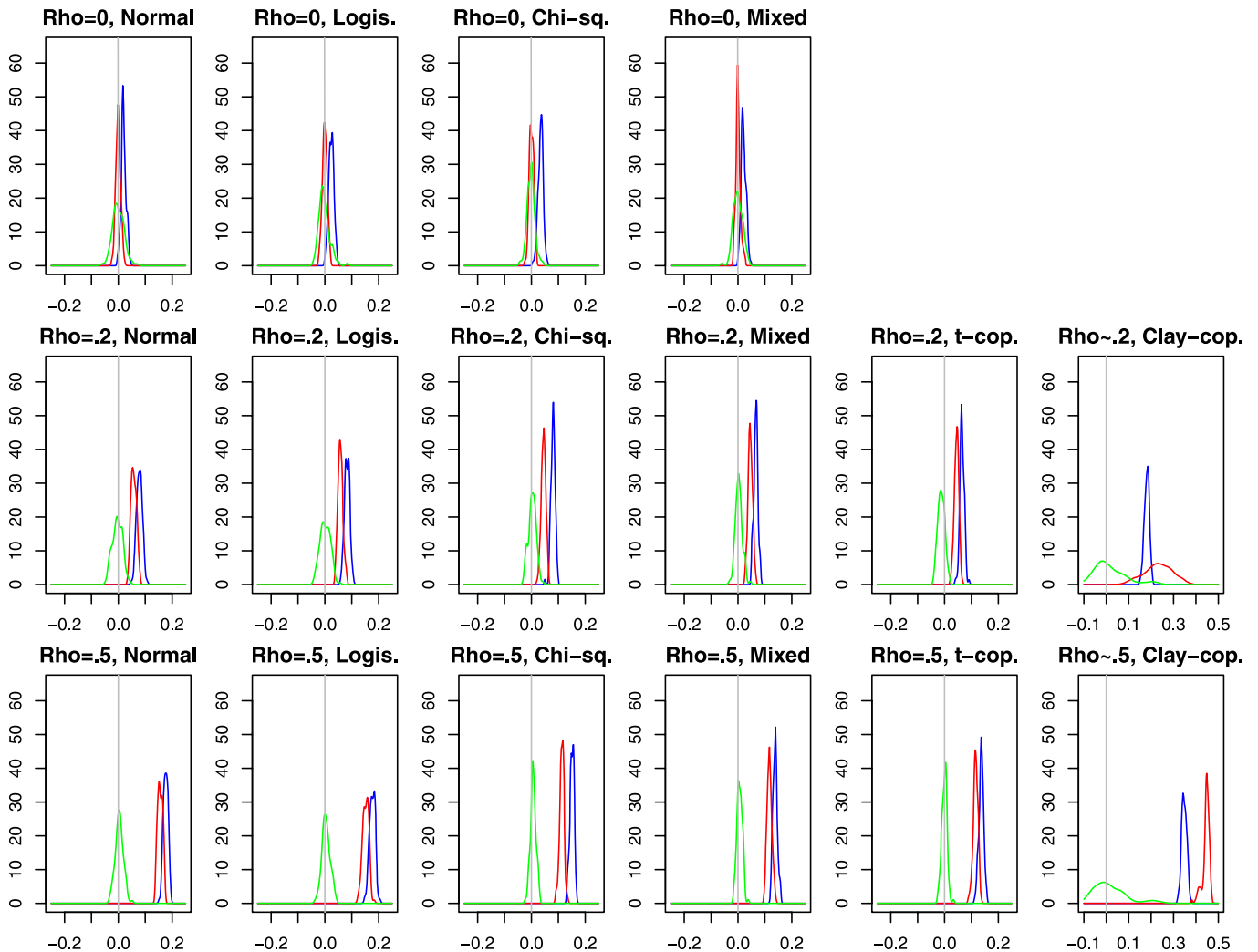


Figure 2. Monte Carlo results: The red (dark grey) density indicates the error of the sample proportion, the blue (medium grey) density indicates the error of the weighted sample proportion, and the green (light grey) density indicates the error of the variable response propensity estimators. ρ is varied across the rows and the model assumptions are varied across the columns. Note that the range of values is different for the plots in the sixth column. The online version of this figure is in color.

When we consider cases 2 through 6 (the various misspecified models), we see that the results are not particularly sensitive to the joint distribution of the error terms. Throughout, similar patterns are observed—when ρ is high, the SP and the WSP exhibit significant bias, and the VRPE exhibits a small amount of bias. In theory, the misspecification should increase the bias of the VRPE, but we do not observe this in sample sizes we consider because the misspecification bias in the VRPE is small relative to the sampling error in the estimates.

It is only in case 6, which represents an extreme departure from bivariate normality, that the VRPE performs poorly. The bias of the estimator (when $\rho \approx 0.5$) is only 1.7%, but the standard deviation of the VRPE minus the FSP is 7.3%, the asymptotic covariance matrix is poorly conditioned, and the standard errors provide a poor representation of the sampling variance of the VRPE. In this case however, selection bias is severe. The SP and WSP are severely biased and the performance of the WSP is especially poor. While case 6 presents a situation where the VRPE performs poorly, it also represents a case where conventional estimators are particularly ineffective. Hence, I argue that the VRPE is reasonably robust to the assumption of bivariate normality.

5. APPLICATION TO THE NATIONAL ELECTION STUDIES

In this section, I consider an application of the variable response propensity estimator. In searching for an application, I considered a number of factors. First, the survey should include as many measures of response propensity as possible. Second, the survey should include quantities that are known from other sources, so that the performance of the method can be evaluated. Third, the survey should suffer from significant nonresponse bias, and some of this bias should persist after weighting by demographics characteristics. The voter turnout items in the 1980, 1984, and 1988 American National Election Studies provides the ideal test case, meeting the conditions outlined above.

5.1 Estimating Voter Turnout

The first step is to use validated voter turnout in the ANES survey to infer the turnout rates in the 1980, 1984, and 1988 U.S. presidential elections. It is well known that self-reported voter turnout in the ANES suffers from misreporting (Katz 2000). Using the validated turnout item allows us to isolate the problem of nonresponse bias from the problem of measurement error. The ANES includes a preelection and a postelection component (using the same respondents). The voter turnout item is reported for respondents who participated in both the preelection and postelection surveys. The total response rates for the postelection survey vary between 61% and 63% for the years we consider. Most nonresponse in the ANES is attributable to refusal to participate in the preelection or postelection surveys, as opposed to noncontact and other sources.

In Table 1, I report the actual turnout rates. I also report simple proportions based on self-reported and validated turnout. Self-reported turnout overestimates actual turnout by between 19% and 20%. Using validated turnout reduces the discrepancy to between 9% and 12%. We can take the validated turnout rates as the starting point because it allows us to isolate the problem of nonresponse bias from the problem of measurement error.

The next step in the correction process is to employ poststratification weights. I construct 32 demographic cells based on race, gender, age, and educational attainment. Race is divided into two categories (black and other), age is divided into four categories (18–29, 30–44, 45–59, and 60+) and educational attainment is divided into three categories (less than high school, high school, and some college). The choice of variables was dictated by two concerns. First, uncontroversial weighting targets should exist and be provided in U.S. Census data. Second, the variable should contain little item nonresponse (which excludes income). To obtain weights for these targets, I relied on demographic data from the U.S. Census.

The weighting targets indeed differed from the values found in the ANES. The most significant differences were that the

Table 1. Estimates of voter turnout

	1980	1984	1988
Actual Voter Turnout	52.8%	53.3%	50.3%
Simple Estimators			
Self-Reported Turnout	71.3% (1.2%)	73.6% (1.0%)	69.6% (1.1%)
Simple Proportion (Validated)	62.0% (1.3%)	64.8% (1.1%)	59.8% (1.2%)
Weighted Proportion (Validated)	59.0% (1.3%)	61.8% (1.1%)	57.7% (1.2%)
Corrected Estimates (Classes)			
Calls	60.2% (1.6%)	62.3% (1.0%)	55.7% (1.2%)
Letter	59.7% (2.2%)	—	55.1% (1.5%)
Cooperation	55.2% (1.8%)	60.3% (1.4%)	54.5% (1.6%)
Interest	53.2% (1.5%)	58.7% (1.3%)	52.1% (1.5%)
Conversion	—	61.8% (2.1%)	57.8% (3.4%)
Corrected Estimates (VRPE)			
Calls	60.4% (3.7%)	59.5% (3.2%)	53.9% (3.3%)
Letter	56.7% (5.0%)	—	52.8% (3.8%)
Cooperation	54.5% (3.8%)	56.2% (3.2%)	51.4% (3.4%)
Interest	52.5% (3.7%)	55.7% (3.1%)	50.9% (3.3%)
Conversion	—	56.9% (4.2%)	58.6% (4.8%)

NOTE: Standard errors are in parentheses.

ANES included a higher proportion of older and highly educated individuals. In Table 1, I report a demographics-weighted proportion. The weighted proportion overestimates turnout by between 6% and 8%. Weighting by demographic characteristics succeeds in reducing nonresponse bias, but it does not eliminate it completely.

More sophisticated methods of correcting for nonresponse bias require measuring response propensity. I considered five different measures—the number of calls or visits to the household, whether a persuasion letter was sent, whether the household initially refused, and interviewer-coded measures of interest in the interview and cooperativeness.

In Figure 3, I graph turnout against three measures of response propensity. For cooperativeness and interest, we see strong monotonic relationships. Low-propensity respondents vote at lower rates indicating that corrected estimates would lead to lower estimates of turnout. These same patterns persisted after I controlled for demographic characteristics. For number of calls, the relationship between turnout and response propensity is much weaker, suggesting that corrections based on the number of calls will lead to smaller corrections.

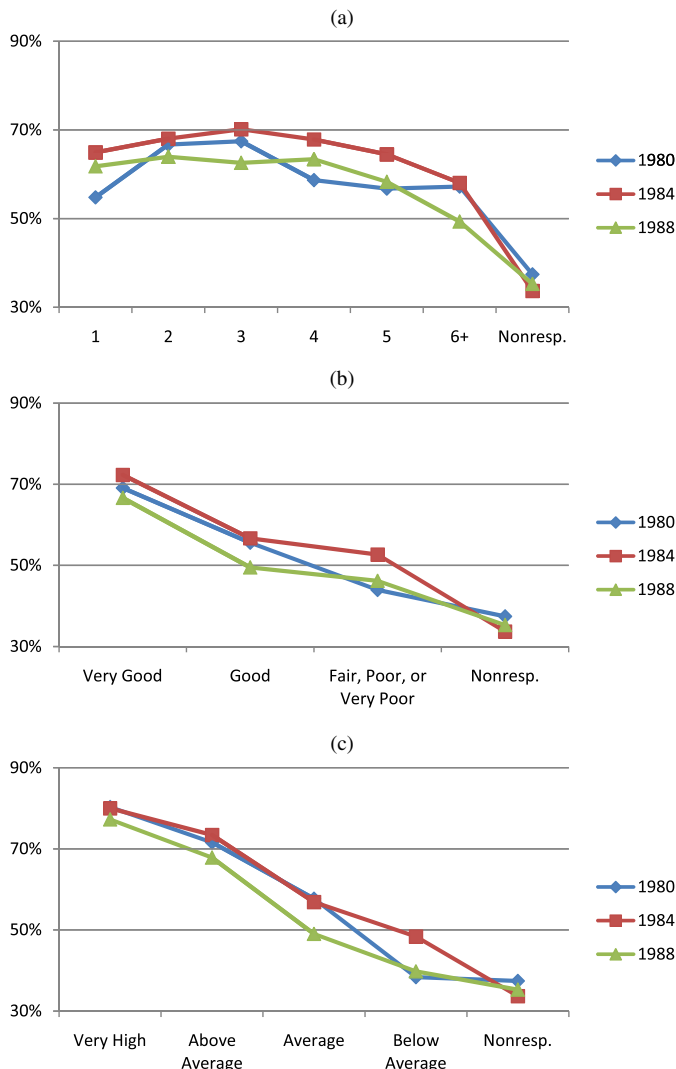


Figure 3. Turnout by response propensity. (a) Number of calls. (b) Cooperation. (c) Interest. The online version of this figure is in color.

I am able to report turnout for the nonrespondents in Figure 3 because the true voter turnout rate is known from election results. Knowing the turnout rate among the respondents, the overall turnout rate, and the response rate is sufficient to determine the turnout rate among the nonrespondents. This figure indicates that extrapolation based on response propensity is likely to be successful. The nonrespondents indeed resemble the low-propensity respondents more than the high-propensity respondents. Hence, the fundamental identifying assumption of the variable response propensity estimator is satisfied in the ANES turnout data.

Before considering the variable response propensity estimator, I consider a simpler correction based on the classes model. One difficulty of applying this method as described in *Stinchcombe, Jones, and Sheatsley (1981)* to the ANES data is that we cannot separate refusals from noncontacts (except in 1988, where an auxiliary nonresponse file is available), but ignoring noncontacts is unlikely to cause problems here because there are so few noncontacted individuals among the nonrespondents. Instead, I implemented the classes correction by assuming that all nonrespondents voted at the same rate as the low-propensity respondents. Table 1 reports estimates based on the classes model, for all five measures of response propensity. These measures led to an improvement over the weighted proportions reported in Table 1, but did not fully correct for nonresponse bias. As Figure 3 indicates, the nonrespondents exhibit lower voter turnout than the low-propensity respondents, while the classes model assumes that they exhibit the same turnout rate as the low-propensity respondents.

Finally, I compute results for the variable response propensity estimator. In the first step, I estimate the parameters of the model using maximum likelihood. I consider a similar specification for latent voting propensity and latent response propensity. I estimate the model

$$y_n^* = \alpha_1 + \alpha_2 Educ2_n + \alpha_3 Educ3_n + \alpha_4 Female_n + \alpha_5 Black_n + \alpha_6 Age2_n + \alpha_7 Age3_n + \alpha_8 Age4_n + \varepsilon_n,$$

$$r_n^* = \beta_1 + \beta_2 Educ2_n + \beta_3 Educ3_n + \beta_4 Female_n + \beta_5 Black_n + \beta_6 Age2_n + \beta_7 Age3_n + \beta_8 Age4_n + \eta_n,$$

where $Educ2_n$ and $Educ3_n$ are dummy variables summarizing education and $Age2_n$, $Age3_n$, and $Age4_n$ are dummy variables summarizing age. The variable response propensity estimator also requires selecting weighting targets. I used the same weighting targets as I did for the demographics-weighted proportion.

For space considerations, I only report the full estimation results for the cooperation measure. These results are reported in Table 2. Looking at the outcome equation, the results indicate that older and more educated respondents were far more likely to vote in the presidential elections I analyze. Blacks were less likely to vote in 1984 and 1988. The coefficients on female are statistically insignificant. Looking at the selection equation, we find that older and more educated respondents are more likely to have been deemed cooperative by the interviewers. This is consistent with my earlier finding, which suggested that weighting by age and education significantly affected the estimates of turnout.

Table 2. Coefficient estimates

	1980	1984	1988
Outcome Equation			
Constant	-0.937*** (0.123)	-0.899*** (0.104)	-0.971*** (0.111)
Female	-0.074 (0.072)	0.096 (0.061)	-0.020 (0.064)
Black	0.006 (0.112)	-0.287 ** (0.097)	-0.401*** (0.097)
Age: 18–29	—	—	—
Age: 30–44	0.628*** (0.093)	0.421*** (0.078)	0.528*** (0.085)
Age: 45–59	0.857*** (0.104)	0.786*** (0.094)	0.789*** (0.100)
Age: 60+	1.015*** (0.109)	0.989*** (0.092)	0.967*** (0.096)
Educ: Less than H.S.	—	—	—
Educ: Graduated H.S.	0.524*** (0.095)	0.464*** (0.082)	0.392*** (0.085)
Educ: Some College +	1.026*** (0.102)	1.075*** (0.088)	1.004*** (0.087)
Nonresponse Equation			
Constant	-0.439*** (0.093)	-0.380*** (0.078)	-0.283*** (0.084)
Female	0.118 (0.067)	0.048 (0.058)	0.115 (0.061)
Black	0.067 (0.089)	-0.007 (0.085)	0.036 (0.084)
Age: 18–29	—	—	—
Age: 30–44	0.344*** (0.089)	0.368*** (0.076)	0.353*** (0.081)
Age: 45–59	0.351*** (0.100)	0.302*** (0.089)	0.256 ** (0.093)
Age: 60+	0.510*** (0.098)	0.394*** (0.082)	0.418*** (0.087)
Educ: Less than H.S.	—	—	—
Educ: Graduated H.S.	0.334*** (0.082)	0.367*** (0.071)	0.084 (0.072)
Educ: Some College +	1.062*** (0.098)	0.953*** (0.077)	0.635*** (0.079)
ρ	0.227*** (0.062)	0.258*** (0.049)	0.265*** (0.056)

NOTE: Standard errors are in parentheses. One star indicates significance at the 5% level, two stars indicates significance at the 1% level, and three stars indicates significance at the 0.1% level.

The parameter ρ captures the correlation between the unobservables in the two equations. The coefficients range between 0.23 and 0.27, and are highly statistically significant in the years I analyze. This indicates that nonignorable selection bias is in fact a serious problem, which must be corrected for.

I report the variable response propensity estimates in Table 1. The results for number of calls are mixed. Although the correction is in the right direction in 1988 (relative to the weighted proportion), the estimator yields a small correction in 1980 and 1984. The results for letter are somewhat better. In both cases, the estimate is substantially closer to the truth. Using cooperation or interest, we almost completely eliminate nonresponse bias. The results for refusal conversion are more mixed, but once again offer an improvement over the weighted proportion.

Relative to the estimates reported for the classes model, the variable response propensity estimator offers an improvement as well. The classes model assumes that nonrespondents resemble low-propensity respondents, while in fact, nonrespondents tend to be more extreme than low-propensity respondents. One may initially think that it provides an “unfair” comparison to employ weighting in the variable response propensity estimator, but not the classes estimator. Employing weights to match the weighting targets in the achieved sample in the classes estimator would lead to an extrapolated sample that no longer matched those weighting targets. Alternatively, applying the classes correction first and then weighting would lead to a weighted proportion of nonrespondents that did not match the actual proportion of nonrespondents. It is perhaps for this reason that a “weighted classes method” does not exist in the literature. I developed a weighted classes method that solved these two

problems and found that the estimates provided an improvement over the classes method, but did not preform as well as the VRPE.

5.2 Comparison to a Selection Model

My framework differs from a sample selection model in that more than one category of response is observed among the respondents. It is this difference that allows me to identify the model in the absence of exclusion restrictions (which I argue are very difficult to obtain). Here, I compare my results to the results of a selection model based on survey administration variables. This follows [Brehm \(1993\)](#)’s approach, but I extend his method to correcting for nonresponse bias in a proportion ([Brehm](#) only considers correcting for nonresponse bias in regression coefficients).

Employing [Brehm](#)’s approach requires access to auxiliary information about the nonrespondents. This information are available in the 1986 and 1988 American National Election studies. I apply [Brehm](#)’s method to these cases and compare the results to the SP, the WSP, and the VRPE. I estimate a selection model using the same variables [Brehm](#) employs. In the second stage, I estimate a linear model including the same covariates as I included in the previous subsection. Here, a one-step estimator tailored to a binary dependent variable may be more appropriate, but I chose to follow [Brehm](#)’s approach for comparability.

The estimate of the sample proportion is given by $\hat{\pi}_3 = \sum_{j=1}^J p_j^x \hat{\alpha}' \tilde{x}_j$, where $\hat{\alpha}$ is the estimate obtained from the outcome equation in the selection model. The choice of covariates was dictated by the necessity of having estimates of the population cells p_j^x (i.e., I could only consider a limited set of covariates for which the distribution of the population is known). This

Table 3. Selection model estimates

	1986		1988	
	Uncorrected	Corrected	Uncorrected	Corrected
Outcome Equation				
Constant	—	1.790*** (0.069)	—	2.084*** (0.083)
Persuasion Letter	—	-2.263*** (0.075)	—	-2.211*** (0.078)
Initial Refusal	—	-0.264 (0.137)	—	-0.298 (0.176)
Log(Calls)	—	-0.164** (0.052)	—	-0.333*** (0.056)
Nonresponse Equation				
Constant	0.028 (0.028)	0.038 (0.028)	0.237*** (0.036)	0.241*** (0.036)
Female	0.020 (0.020)	0.021 (0.020)	-0.020 (0.022)	-0.023 (0.022)
Black	-0.053 (0.028)	-0.054 (0.028)	-0.158*** (0.035)	-0.155*** (0.035)
Age: 18–29	—	—	—	—
Age: 30–44	0.146*** (0.025)	0.144*** (0.025)	0.175*** (0.030)	0.180*** (0.030)
Age: 45–59	0.327*** (0.030)	0.326*** (0.030)	0.276*** (0.034)	0.275*** (0.034)
Age: 60+	0.465*** (0.028)	0.464*** (0.029)	0.325*** (0.032)	0.331*** (0.032)
Educ: Less than H.S.	—	—	—	—
Educ: Graduated H.S.	0.177*** (0.026)	0.180*** (0.026)	0.325*** (0.032)	0.153*** (0.031)
Educ: Some College +	0.326*** (0.025)	0.327*** (0.025)	0.146*** (0.031)	0.342*** (0.029)
$\rho\sigma$	—	-0.047 (0.029)	—	-0.049 (0.034)
Actual Voter Turnout	36.4%		50.3%	
Estimates				
Sample Proportion	44.7% (0.8%)		59.8% (1.2%)	
Weighted Sample Proportion	43.1% (0.8%)		57.7% (1.2%)	
VRPE	38.1% (2.8%)		51.4% (3.4%)	
Selection Model	43.0% (1.0%)	44.1% (1.2%)	57.3% (1.1%)	58.2% (1.3%)

NOTE: Standard errors are in parentheses. One star indicates significance at the 5% level, two stars indicates significance at the 1% level, and three stars indicates significance at the 0.1% level.

presents one difference with Brehm’s study—he was free to select a more extensive set of covariates in the outcome equation because he did not have to know their distribution in the population.

The results are given in Table 3. As can be seen, the estimate of ρ is small in magnitude and not statistically significantly different from zero. Taken literally, this estimate suggests that nonignorable nonresponse is not present (we, of course, know that this is not the case). Consequently, the selection model estimates are close to the WSP and far from the true values. For both 1986 and 1988, the VRPE provide a substantial improvement over the SP and the WSP, and nearly eliminates nonresponse bias.

5.3 Nonresponse Bias in Other Quantities

Previously, I demonstrated that the estimator I proposed is able to diagnose and correct for nonresponse bias. Here, I apply the same estimator to diagnose and correct for nonresponse bias in a number of items in the ANES in the same years studied earlier. I included both items expected to relate to interest in politics (e.g., whether an individual read about the presidential campaign in a magazine) and ideology (whether an individual supports a woman’s right to obtain an abortion).

By its’ nature of being a survey of political behavior, participation in the ANES should be positively related to interest in politics. Hence, we would expect that weighted proportions would overestimate (underestimate) those items that are positively (negatively) associated with interest in politics. Furthermore, while both liberal and conservative talking heads find it

convenient to dismiss polls that are not favorable to their cause, conservatives are often harsher in their criticisms of polls and more distrustful of the media organizations that frequently conduct public opinion polls. If the attitudes of conservative elites influence the general attitudes of conservatives towards polling organizations, then we should expect conservatives to be more reluctant to participate in political polls. Thus, we would expect that weighted proportions will provide an underestimate (overestimate) for those items that are positively (negatively) associated with conservatism. Items closely associated with interest in politics should, however, exhibit relatively more bias than items related to ideology.

I report results in Table 4. For each item, I report the WSP and the VRPE, using interviewer-coded cooperativeness as the measure of response propensity. The results here are consistent with expectations. Items closely related to interest in or knowledge of politics (percent who think there are important differences between the parties, listen to campaign speeches on the radio, read about the campaign in a magazine, and think that there are quite a few crooks in government) usually indicate that nonresponse bias is present. The bias ranges from 1.1% to 7.2%. Items closely related to ideology (prochoice attitudes and antischool prayer attitudes) exhibit small biases and are statistically significant in only one out of five cases. The bias ranges from 0.7% to 1.8%. The items involving identification with the Democratic and Republican parties are potentially related to both interest and politics and ideology. Here, the results suggest that the weighted proportion provides an accurate estimate of the proportion of Democratic identifiers, underestimates the

Table 4. Weighted proportion and variable response propensity estimates

Percent of American adults that	1980			1984			1988		
	WSP	VRPE	Diff.	WSP	VRPE	Diff.	WSP	VRPE	Diff.
... think there are any important differences in what the Republicans and Democrats stand for	55.8%	49.0%	6.8%*				58.0%	50.8%	7.2%*
... identify with the Democratic Party	38.7%	38.5%	0.2%	35.9%	35.2%	0.7%	34.4%	34.5%	-0.2%
... identify with the Republican Party	22.5%	21.3%	1.2%	26.9%	25.3%	1.6%*	27.8%	25.7%	2.2%*
... listen to any speeches or discussions about the presidential campaign on the radio	46.4%	43.5%	2.9%*	43.9%	40.7%	3.2%*	31.1%	28.9%	2.1%
... read about the presidential campaign in any magazines	32.4%	29.0%	3.3%*	31.8%	28.2%	3.6%*	24.3%	20.5%	3.8%*
... think that quite a few of the people running the government are crooked	46.9%	48.5%	-1.6%	31.3%	30.6%	0.7%			
... view that, by law, a woman should always be able to obtain an abortion as a matter of personal choice	35.1%	33.2%	1.8%*	34.4%	33.2%	1.1%	35.4%	33.8%	1.6%
... thinks that religion does not belong in the school	20.4%	19.4%	1.0%	20.4%	19.7%	0.7%			

NOTE: Standard errors are in parentheses. One star indicates significance at the 5% level, two stars indicates significance at the 1% level, and three stars indicates significance at the 0.1% level.

proportion of independents, and overestimates the proportion of Republican identifiers. These differences do not attain statistical significance in 1980, but we note that the results in 1980 are not statistically significantly different than the estimates found in 1984 and 1988.

I note that the VRPE typically has a larger standard error than the WSP. This is not surprising and suggests an efficiency loss when applying the VRPE when the data are missing at random (a similar result was found in the Monte Carlo experiments). Hence, even when all the necessary assumptions hold, there is a bias-variance tradeoff in applying the VRPE. For this reason, I would recommend a pretesting approach. In the event that ρ is statistically indistinguishable from zero, there is no evidence of nonignorable nonresponse bias and the WSP proportion should be reported. It is only for items for which $\rho = 0$ can be rejected that I would recommend reporting the VRPE. It is in these items where the WSP is likely to exhibit substantial bias.

6. APPLICATION TO PUBLIC OPINION POLLING

I considered a second application of the variable response propensity estimator—this time to a public opinion poll conducted in September, 2004, sponsored by CBS News and the New York Times. The poll includes a measure of response

propensity—respondents were asked whether they would be willing to have a reporter call them back in a few days to discuss their views further. Individuals who responded that they were willing were further coded by the interviewer as being “talkative” or “not talkative.” This item closely resembles the measure I found to be most effective in the previous section. Perhaps the only drawback of this question is that respondent were asked whether they would be willing to talk to a *reporter*, which may prompt negative feelings among conservatives, who are exposed to media that (ironically) is frequently critical of the media.

I considered a number of different variables including approval of President Bush, whether the respondent voted in the 2002 midterm elections, whether the respondent believed that the Iraq war was worth the cost, and whether the respondent believed that Saddam Hussein was not involved in the 9/11 terror attacks. In Table 5, I report the estimate of ρ from the variable response propensity model and I report the SP, the WSP, and the VRPE, for these items.

For two of the items (the respondent believes that Saddam Hussein was not involved in 9/11 and the respondent reported voting in the 2002 midterm elections), we can reject $\rho = 0$ at the 5% level. We find that the WSP overestimates both propor-

Table 5. Estimates for the September, 2004 CBS/NYT preelection poll

Percentage of American adults who that	ρ	SP	WSP	VRPE
... approve of President Bush	-0.058 (0.061)	52.1% (1.4%)	51.3% (1.4%)	52.7% (5.0%)
... believe the Iraq war was worth the cost	0.093 (0.056)	40.9% (1.4%)	38.4% (1.4%)	34.8% (4.5%)
... believe that Saddam Hussein was not involved in 9/11 terror attacks	0.153** (0.058)	51.3% (1.4%)	47.5% (1.4%)	41.0% (4.7%)
... report voting in 2002	0.195*** (0.057)	55.8% (1.4%)	46.6% (1.4%)	39.3% (4.1%)

NOTE: Standard errors are in parentheses. One star indicates significance at the 5% level, two stars indicates significance at the 1% level, and three stars indicates significance at the 0.1% level.

tions. For a third item (the respondent believes that Iraq war was worth the cost), we can reject $\rho = 0$ at the 10% level. In this case, the VRPE results in a more modest correction to the WSP. We do not find any evidence of nonignorable nonresponse bias in the estimate of President Bush's approval rating, and the resulting correction is small in magnitude.

As we found with the ANES, the item relating to political participation (voting in the 2002 midterm election) exhibits significant nonresponse bias. The VRPE is substantially closer to the actual voter turnout rate (which was 37.0%) than the WSP. I note however that this item should be interpreted as an estimate of the proportion of Americans who *report* voting in the 2002 midterm election, which may be different than the proportion that actually voted due to misreporting.

Whether the respondent believes that Saddam Hussein was not involved in the 9/11 terror attacks also exhibited significant bias. Individuals who are politically informed may be more likely to agree to participate in a preelection poll conducted by a news organization, and may also be more likely to hold the view that Saddam Hussein was not involved in the 9/11 attacks.

With more information, a more detailed analysis could be conducted. The documentation for the poll did not separate noncontacts from refusals. Moreover, the poll did not report the number of callbacks, which would be useful in diagnosing noncontact bias. Ideally, this information would be present.

7. DISCUSSION

While poststratification weights are widely employed for correcting survey estimates, methods for correcting for nonignorable nonresponse bias are far less prevalent. I derived a new estimator that is capable of correcting for nonignorable nonresponse bias while retaining the advantages of post-stratification on demographic characteristics. I used the estimator to correct for unit nonresponse bias in the American National Election Studies. The results indicate that the estimator performs quite well. Estimates of voter turnout from the ANES are severely biased, even after measurement error is corrected for using the voter validation studies. My method never leads to worse estimates than the demographics-weighted average, and usually provides a substantial improvement.

I found that interviewer-coded measures of response propensity were the most successful, while the number of calls was the least successful measure. This result is not altogether surprising—most of the nonresponse in the ANES is caused by refusal, rather than failure to contact. Hence, nonrespondents are better characterized as similar to uncooperative respondents rather than hard to locate respondents.

My results can potentially be applied to correct for nonresponse bias in a large range of surveys. Without extensive experience with similar variables in similar surveys indicating otherwise, there is *always* a potential that the items of interest exhibit substantial nonresponse bias. My results indicate that items closely related to interest in politics are likely to exhibit substantial nonresponse bias, even in the “gold standard” of political surveys. Most public opinion polls obtain far lower response rates than the ANES, indicating that nonresponse bias is likely to be even more severe. Fortunately, my method provides a way to deal with this problem.

A limitation of my findings is that academic surveys such as the ANES present somewhat of an easy case, in comparison to most public opinion polls. The ANES preselects respondents for inclusion in the sample, and the survey is conducted over a long time period, minimizing the number of noncontacts. In contrast, most public opinion polls are less systematic about who enters the sample. Rather than preselecting a set of individuals and attempting to contact all of them, individuals are continually selected as the survey proceeds. All contact stops once a target number of respondents are reached. As a result, different levels of effort were made to contact different selected individuals.

A second difference is that while the ANES contacts the vast majority of selected individuals, polls with shorter field period or more limited budgets may see many noncontacts among the nonrespondents. Moreover, there is a potential that noncontact bias and refusal bias may be caused by different mechanisms (Groves and Couper 1998) and may work in different directions (Fitzgerald and Fuller 1982; Lin and Schaeffer 1995; Voigt, Koepsell, and Daling 2003). In many applications, the lack of a pattern in the number of callbacks may itself indicate that noncontact bias is minimal. In these cases, it may be sufficient to extrapolate from the low-propensity respondents to the refusals, and ignoring the noncontacts. When this is not the case (when both refusal and noncontact bias are present) applications of the VRPE may consider incorporating two equations—a refusal equation and a contact equation—into the survey response mechanism.

Finally, I note that public opinion polls are designed with two concerns in mind—maximizing response rates and minimizing cost. The literature has found that marginal increases in response rates do not seem to reduce nonresponse bias. My results suggest that the cost-benefit analysis should be reconsidered. Survey researchers will obtain more accurate estimates by shifting the focus from marginal increases in response rates to constructing the sample in a more systematic way. This will allow survey researchers to better diagnose nonresponse bias and correct for it. However, even when surveys are conducted in a relatively un-systemic way (this includes public opinion polls conducted over short periods of time and internet surveys that do not rely on a probability sample), I believe that response propensity is a valuable diagnostic for nonresponse bias.

[Received August 2009. Revised August 2010.]

REFERENCES

- Arabmazar, A., and Schmidt, P. (1982), “An Investigation of the Robustness of the Tobit Estimator to Nonnormality,” *Econometrica*, 50, 1055–1063. [5]
- Bethlehem, J. J. (1988), “Reduction of Nonresponse Bias Through Regression Estimation,” *Journal of Official Statistics*, 4, 251–260. [4]
- Biemer, P. P., and Link, M. W. (2008), “Evaluating and Modeling Early Cooperator Effects in RDD Surveys,” in *Advances in Telephone Survey Methodology*, Hoboken, NJ: Wiley. [2,3]
- Brehm, J. (1993), *The Phantom Respondents*, Ann Arbor: University of Michigan Press. [2,3,9,10]
- (1999), “Alternative Corrections for Sample Truncation: Application to the 1988, 1990, and 1992 Senate Election Studies,” *Political Analysis*, 8, 147–165. [3]
- Brick, J. M., Martin, D., Warren, P., and Wivagg, J. (2003), “Increased Efforts in RDD Surveys,” in *2003 Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association. [1]
- Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., and Chapman, C. (2005), “Implications for RDD Design From an Incentive Experiment,” *Journal of Official Statistics*, 21, 571–589. [1]

- Chamberlain, G. (1986), "Asymptotic Efficiency in Semi-Parametric Models With Censoring," *Journal of Econometrics*, 32, 189–218. [5]
- Curtin, R., Presser, S., and Singer, E. (2000), "The Effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly*, 64, 413–428. [1,2]
- Drew, J. H., and Fuller, W. A. (1980), "Modeling Nonresponse in Surveys With Callbacks," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 639–642. [2,3]
- (1981), "Nonresponse in Complex Multiphase Surveys," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 623–628. [2,3]
- Dunkelberg, W. C., and Day, G. S. (1973), "Nonresponse Bias and Callbacks in Sample Surveys," *Journal of Marketing Research*, 10, 160–168. [2]
- Ellis, R. A., Endo, C. M., and Armer, J. M. (1970), "The Use of Potential Nonrespondents for Studying, Nonresponse Bias," *Pacific Sociological Review*, 13, 103–109. [2,3]
- Filion, F. L. (1975), "Estimating Bias Due to Nonresponse in Mail Surveys," *Public Opinion Quarterly*, 39, 482–491. [2,3]
- (1976), "Exploring and Correcting for Nonresponse Bias Using Follow-Ups of Nonrespondents," *Pacific Sociological Review*, 12, 103–109. [2,3]
- Fitzgerald, R., and Fuller, L. (1982), "I Hear You Knocking but You Cant Come In: The Effects of Reluctant Respondents and Refusers and Sample Survey Estimates," *Sociological Methods and Research*, 11, 3–32. [2,12]
- Geweke, J. F., Keane, M. P., and Runkle, D. E. (1994), "Alternative Computational Approaches to Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76, 609–632. [4]
- Goldberger, A. (1983), "Abnormal Selection Bias," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, New York: Academic Press. [5]
- Green, W. H. (2000), *Econometric Analysis* (4th ed.), Upper Saddle River, NJ: Prentice Hall. [4]
- Groves, R. M., and Couper, M. P. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley. [12]
- Groves, R. M., and Peytcheva, E. (2008), "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis," *Public Opinion Quarterly*, 72, 167–189. [2]
- Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. (2002), *Survey Nonresponse*, New York: Wiley. [2]
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004), *Survey Methodology*, New York: Wiley. [2]
- Groves, R. M., Presser, S., and Dipko, S. (2004), "The Role of Topic Interest in Survey Participation Decisions," *Public Opinion Quarterly*, 68, 2–31. [2]
- Hawkins, D. F. (1975), "Estimation of Nonresponse Bias," *Sociological Methods and Research*, 3, 461–485. [2]
- Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–162. [3]
- Katz, J. N. (2000), "Correcting for Survey Misreports Using Auxiliary Information," working paper, California Institute of Technology. [7]
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., and Presser, S. (2000), "Consequences of Reducing Nonresponse in a National Telephone Survey," *Public Opinion Quarterly*, 2, 125–148. [1,2]
- Lin, I.-F., and Schaeffer, N. C. (1995), "Using Survey Participants to Estimate the Impact of Nonparticipation," *Public Opinion Quarterly*, 2, 236–258. [2,3,12]
- Little, J. R. (1986), "Survey Nonresponse Adjustments for Estimates of Means," *International Statistical Review*, 54, 139–157. [4]
- Montaquila, J. M., Brick, J. M., Hagedorn, M. C., Kennedy, C., and Keeter, S. (2008), "Aspects of Nonresponse Bias in RDD Telephone Surveys," in *Advances in Telephone Survey Methodology*, Hoboken, NJ: Wiley. [2]
- O'Neil, M. J. (1979), "Estimating Nonresponse Bias Due to Refusals in Telephone Surveys," *Public Opinion Quarterly*, 43, 218–232. [2]
- Potthoff, R. F., Manton, K. G., and Woodbury, M. A. (1993), "Correcting for Nonavailability Bias in Surveys by Weighting Based on Number of Callbacks," *Journal of the American Statistical Association*, 88, 1197–1207. [2,3]
- Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies of Causal Effects," *Biometrika*, 70, 41–55. [4]
- Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley. [1]
- Smith, T. W. (1984), "Estimating Nonresponse Bias With Temporary Refusals," *Sociological Perspectives*, 27, 473–489. [2]
- Stinchcombe, A. L., Jones, C., and Sheatsley, P. (1981), "Nonresponse Bias for Attitude Questions," *Public Opinion Quarterly*, 45, 359–375. [2,8]
- Teitler, J. O., Reichman, N. E., and Sprachman, S. (2003), "Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population," *Public Opinion Quarterly*, 67, 126–138. [2,3]
- Traugott, M. W. (1987), "The Importance of Persistence in Respondent Selection for Preelection Surveys," *Public Opinion Quarterly*, 51, 48–57. [2]
- Voigt, L. F., Koepsell, T. D., and Daling, J. R. (2003), "Characteristics of Telephone Survey Respondents According to Willingness to Participate," *American Journal of Epidemiology*, 157, 66–73. [2,12]
- Voss, D. S., Gelman, A., and King, G. (1995), "Preelection Survey Methodology: Details From Eight Polling Organizations, 1988 and 1992," *Public Opinion Quarterly*, 59, 98–132. [2]