Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Imputation for nonmonotone last-value-dependent nonrespondents in longitudinal surveys

by Jing Xu, Jun Shao, Mari Palta and Lin Wang

December 2008

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2008
Vol. 34, No. 2, pp. 153-162
Statistics Canada, Catalogue No. 12-001-X

Imputation for nonmonotone last-value-dependent nonrespondents in longitudinal surveys

Jing Xu, Jun Shao, Mari Palta and Lin Wang [1]

Abstract

In longitudinal surveys nonresponse often occurs in a pattern that is not monotone. We consider estimation of time-dependent means under the assumption that the nonresponse mechanism is last-value-dependent. Since the last value itself may be missing when nonresponse is nonmonotone, the nonresponse mechanism under consideration is nonignorable. We propose an imputation method by first deriving some regression imputation models according to the nonresponse mechanism and then applying nonparametric regression imputation. We assume that the longitudinal data follow a Markov chain with finite second-order moments. No other assumption is imposed on the joint distribution of longitudinal data and their nonresponse indicators. A bootstrap method is applied for variance estimation. Some simulation results and an example concerning the Current Employment Survey are presented.

Key Words: Bootstrap; Nonmonotone missingness; Last-value-dependent; Nonignorable nonresponse; Nonparametric regression.

1. Introduction

A survey is longitudinal if data are collected from every sampled unit at multiple time points. For example, in the Current Employment Survey (CES), commonly known as the payroll survey conducted by the U.S. Bureau of Labor Statistics, data are obtained from establishments on a monthly basis by mail, telephone, FAX, and electronic data entry (Butani, Harter and Wolter 1997). Other examples include the Survey of Income and Program Participation (SIPP) and many economic surveys conducted by the U.S. Census Bureau. Nonresponse occurs in longitudinal studies. We assume that every sampled unit responds at baseline (the first time point). Nonresponse is monotone if a unit not responding at some time does not return to the survey. Nonmonotone nonresponse, however, often occurs in surveys such as the CES and SIPP and entails a wider variety of nonresponse patterns.

Let $y_1, \ldots, y_T$ be the values of a variable from a sample unit, where T is the total number of time points, and $\delta_1, \ldots, \delta_T$ be the response indicators ($\delta_t = 1$ if y_t is a respondent and $\delta_t = 0$ if y_t is a nonrespondent). Nonresponse is completely at random if $(\delta_1, \ldots, \delta_T)$ is statistically independent of $(y_1, \ldots, y_T)$, which rarely occurs in surveys. A more realistic assumption is that nonresponse at time point t depends on observed or unobserved past values $y_1, \ldots, y_{t-1}$. In this paper, we focus on a stronger assumption, the last-value-dependent nonresponse mechanism, *i.e.*, nonresponse of y_t depends on the last value y_{t-1} (observed or unobserved). The last-value-dependent nonresponse mechanism is assumed in many economic surveys (*e.g.*, the CES; see Butani, Harter and Wolter 1997). If nonresponse is also monotone, then either y_t is a nonrespondent with certainty or y_{t-1} is observed. This is a special case of what is referred to as ignorable missingness (Little and Rubin 1987). For nonmonotone nonresponse, however, last-value-dependent nonresponse is nonignorable, as whether y_t is a respondent depends on y_{t-1} that may be a nonrespondent.

Existing methods for handling nonmonotone nonresponse can be briefly described as follows. Under parametric modeling, methods such as the maximum likelihood, multiple imputation, or Bayesian analysis can be applied (*e.g.*, Troxel, Harrington and Lipsitz 1998; Troxel, Lipsitz and Harrington 1998; Schafer 1997), if a suitable parametric model for the joint distribution of $(y_1, \ldots, y_T)$ and $(\delta_1, \ldots, \delta_T)$ can be found. The validity of these methods, however, depends on whether parametric models are correctly specified. A simple linear regression imputation method (see, *e.g.*, Butani *et al.* 1997) imputes a nonrespondent y_t by the predicted value under a fitted linear regression model between y_t and y_{t-1}, where the regression model is fitted using data from sampled units with both y_t and y_{t-1} observed and the prediction is made using the predictor being either the observed y_{t-1} or a previously imputed value of a nonrespondent y_{t-1}. Under the nonmonotone nonresponse mechanism (1), however, it can be shown that simple linear regression imputation is biased even if the linear regression model between y_t and y_{t-1} is correct. The bias is mainly caused by the erroneous way of using imputed y_{t-1} values to impute missing y_t values. A censoring approach creates a dataset with monotone nonresponse by discarding all respondents from a sampled unit after its first missing y-value. Methods

1. Jing Xu, Jun Shao and Lin Wang, Department of Statistics, University of Wisconsin, Madison, WI 53706; Mari Palta, Department of Population Health Sciences and Department of Biostatistics and Medical Informatives, University of Wisconsin, Madison, WI 53706.

appropriate for monotone nonresponse (Paik 1997; Robins, Rotnitzky and Zhao 1995; Troxel, Lipsitz and Brennam 1997) can then be applied to the reduced dataset. Although this approach produces correct estimators, it is not efficient when T is not small, since many respondents are discarded.

The purpose of this article is to propose an imputation method for longitudinal surveys with nonmonotone nonresponse and the last-value-dependent nonresponse mechanism (1). Imputation is commonly used to compensate for nonresponse in survey problems (Kalton and Kasprzyk 1986; Rubin 1987). Once all nonrespondents are imputed, estimates of parameters (such as the mean of y_t) are computed using standard methods by treating imputed values as observations. Our proposed imputation method produces approximately unbiased and consistent estimators for the means of $y_1, \ldots, y_T$.

The rest of this paper is organized as follows. In Section 2, we describe our assumptions. Section 3 describes the imputation process. Some properties of the resulting estimators of population means are discussed in Section 4, together with the proposal of a bootstrap procedure for variance estimation. Section 5 contains some simulation results. An example related to the CES is presented in Section 6. The last section contains a summary.

2. Assumption and imputation model

Let P be a finite population indexed by $i = 1, \ldots, N,$ and let S be a sample of size n taken from P according to some sampling design. According to the sampling design, survey weights $w_i, i \in S,$ are constructed so that for any set of values $\{z_i : i \in P\},$

$$E_s\left(\sum_{i \in S} w_i z_i\right) = \sum_{i=1}^{N} z_i,$$

where E_s is the expectation with respect to S. For each unit $i \in P, (y_{1,i}, \ldots, y_{T,i})$ is a vector of items of interest obtained at time points $t = 1, \ldots, T.$ When nonresponse is present, each unit also has the vector $(\delta_{1,i}, \ldots, \delta_{T,i})$ of response indicators. For simplicity, we may omit the index i in our discussion.

We adopt a model-assisted approach by assuming that the vector $(y_{1,i}, \ldots, y_{T,i}, \delta_{1,i}, \ldots, \delta_{T,i})$'s are independent and identically distributed (i.i.d.) from a superpopulation. The i.i.d. assumption can be relaxed by dividing P into several sub-populations (called imputation classes) so that the i.i.d. assumption approximately holds within each imputation class. Imputation classes are usually constructed using a categorical variable whose values are observed for all sampled units; for example, under stratified sampling, strata or unions of strata are often used as imputation classes. Each imputation class should contain a large number of sampled units. When there are many strata of small sizes, imputation classes are often obtained through poststratification (Valliant 1993) and/or combining small strata.

Once imputation classes are constructed, imputation is done within each imputation class. Thus, for simplicity, from now on we assume that there is only one imputation class.

Under the last-value-dependent nonresponse mechanism,

$$P(\delta_t = 1 \mid y_1, \ldots, y_T, \delta_1, \ldots, \delta_{t-1}, \delta_{t+1}, \ldots, \delta_T) = P(\delta_t = 1 \mid y_{t-1}), \quad t = 2, \ldots, T. \quad (1)$$

We do not make any other assumption on $P(\delta_t = 1 | y_{t-1}).$ When there is no nonresponse, we assume that $(y_1, ..., y_T)$ is a Markov chain, *i.e.*,

$$L(y_t \mid y_1, \ldots, y_T) = L(y_t | y_{t-1}), \quad t = 2, \ldots, T, \quad (2)$$

where $L(\xi | \zeta)$ denotes the conditional distribution of ξ given ζ. We do not make any other assumption on $L(y_t \mid y_{t-1})$ except that y_t has a finite second-order moment. In many economic surveys, the following assumption stronger than (2) is assumed:

$$y_t = \beta y_{t-1} + \sqrt{| y_{t-1} |}\, \varepsilon_t, \quad t = 2, \ldots, T, \quad (3)$$

where β is an unknown parameter, ε_t's are independent of y_t's, $\varepsilon_1 = 0,$ and $\varepsilon_2, \ldots, \varepsilon_T$ have mean 0 and a common variance (*e.g.*, the CES data; see Butani *et al.* 1997). Under (3), the best linear unbiased estimator of β is the well known ratio estimator.

To consider asymptotics, we adopt the frame work in Krewski and Rao (1981) and Bickel and Freedman (1984). We assume that the finite population P is a member of a sequence of finite populations indexed by v. All limiting processes are understood to be as $v \to \infty.$ As $v \to \infty,$ the population size N and the sample size n increase to infinity. In sample surveys, the following regularity conditions on w_i's are typically imposed:

$$n \max_{i \in P} w_i \le b_0 N \qquad \text{and} \qquad n \mathrm{Var}_s\left(\sum_{i \in S} w_i\right) \le b_1 N^2, \quad (4)$$

where b_0 and b_1 are some positive constants and Var_s is the variance with respect to sampling. The first condition in (4) ensures that none of the weights w_i is disproportionately large (see Krewski and Rao 1981). The second condition in (4) means that $\mathrm{Var}_s(\sum_{i \in S} w_i / N)$ is at most of the order n^{-1}. Conditions in (4) are satisfied for stratified simple random sampling designs.

3. Imputation process

Our proposed imputation is a type of regression imputation. Thus, one of the key issues to our method is to find the right "predictors" for nonrespondents. For a nonrespondent $y_t, t \ge 2,$ let r be the time point at which the unit has the last observed value, *i.e.*, y_r is observed but $y_{r+1}, ..., y_{t-1}, y_t$ are nonrespondents. Under assumptions (1)-(2), we can use y_r as a predictor in imputing y_t.

3.1 The case of $r = t - 1$

We first consider the case of $r = t - 1$. Let

$$\phi_{t,t-1}(y_{t-1}) = E(y_t | y_{t-1}, \delta_t = 0, \delta_{t-1} = 1)$$

be the conditional expectation (regression function) for a nonrespondent y_t with observed y_{t-1}. If $\phi_{t,t-1}$ is known, we can simply impute y_t by $\phi_{t,t-1}(y_{t-1})$. But $\phi_{t,t-1}$ is usually unknown. It is shown in the Appendix that assumption (1) implies that

$$\phi_{t,t-1}(y_{t-1}) = E(y_t | y_{t-1}, \delta_t = 1, \delta_{t-1} = 1), \quad t = 2, \ldots, T. \quad (5)$$

Thus, $\phi_{t,t-1}$ can be estimated by regressing y_t on y_{t-1} using data from all sampled units having observed y_t and y_{t-1}.

The idea of using (5) for imputation is the same as that in the monotone nonresponse case treated by Paik (1997). Unlike the monotone nonresponse case, however, $\phi_{t,t-1}(x)$ may not be linear in x for nonmonotone nonresponse. Hence, we consider the nonparametric method in Cheng (1994) for regression. The kernel estimator of $\phi_{t,t-1}(x)$ is

$$\hat{\phi}_{t,t-1}(x) = \sum_{i \in S} \kappa\left(\frac{x - y_{t-1,i}}{h}\right) w_i I_{t,t-1,i} y_{t,i} \Bigg/ \sum_{i \in S} \kappa\left(\frac{x - y_{t-1,i}}{h}\right) w_i I_{t,t-1,i},$$

where $\kappa(x)$ is a probability density function, $h > 0$ is a bandwidth, and

$$I_{t,t-1,i} = \begin{cases} 1 & \delta_{t,i} = 1, \delta_{t-1,i} = 1 \quad t = 2, \ldots, T. \\ 0 & \text{otherwise,} \end{cases}$$

A nonrespondent $y_{t,j}$ with respondent $y_{t-1,j}$ is imputed by

$$\tilde{y}_{t,j} = \hat{\phi}_{t,t-1}(y_{t-1,j}).$$

Cheng (1994) suggested a bandwidth $h = Cn^{-2/5}$, where C is a constant. In general, C may be different from application to application, and should be chosen using techniques developed in the kernel estimation literature (*e.g.*, Cheng 1994 and Chapter 5 of Härdle 1990) and/or empirical studies.

3.2 The case of $r < t - 1$

When $r < t - 1$, the situation is more complicated. Let

$$\phi_{t,r}(y_r) = E(y_t | y_r, \delta_t = \cdots = \delta_{r+1} = 0, \delta_r = 1).$$

As nonresponse mechanism (1) is nonignorable, the expected value of y_t conditional on y_r with $r < t - 1$ is not equal for observed and missing y_t, which precludes the use of observed y_t values as outcomes in estimating $\phi_{t,r}$. It is explicitly shown by Xu (2007) that

$$\phi_{t,r}(y_r) \neq E(y_t | y_r, \delta_t = 1, \delta_{t-1} = a_{t-1}, \ldots, \delta_{r+1} = a_{r+1}, \delta_r = 1) \quad (6)$$

where $a_j = 0$ or 1, $j = r+1, \ldots, t-1$. On the contrary, in the case of monotone nonresponse the two sides of (6) are the same (Paik 1997) so that the right hand side of (6) can be used as the regression imputation model and observed y_t values can be used to estimate $\phi_{t,r}$.

We have to find a conditional expectation of y_t (given y_r and some response status) that is equal to $\phi_{t,r}(y_r)$ and enables us to carry out imputation. It is shown in the Appendix that

$$\phi_{t,r}(y_r) = E(y_t | y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1), \quad r = 1, \ldots, t-2, t = 2, \ldots, T. \quad (7)$$

To estimate $\phi_{t,r}$ by fitting regression according to (7), we do not have observed y_t values as outcomes in regression, because units defined by the right hand side of the equation in (7) have $\delta_t = 0$. If we carry out imputation sequentially as $r = t-1, t-2, \ldots, 1$, then the y_t nonrespondents for units with $\delta_{r+1} = 1$ have already been imputed. Thus, we can use these previously imputed y_t values as outcomes in regression. Although at each fixed time point *t*, imputation is carried out sequentially as $r = t-1, t-2, \ldots, 1$, imputation for different time points can be carried out at any order, because at any time point *t*, imputed values are used as outcomes at time point *t* only.

Since $\phi_{t,r}$ is usually not linear, we use the kernel regression. For $t = 2, \ldots, T$ and $r = t-2, t-3, \ldots, 1$, the conditional expectation $\phi_{t,r}(x)$ for any x is estimated by

$$\hat{\phi}_{t,r}(x) = \sum_{i \in S} \kappa\left(\frac{x - y_{r,i}}{h}\right) w_i I_{t,r,i} \tilde{y}_{t,i} \Bigg/ \sum_{i \in S} \kappa\left(\frac{x - y_{r,i}}{h}\right) w_i I_{t,r,i}, \quad (8)$$

where $\tilde{y}_{t,i}$ is a previously imputed value and

$$I_{t,r,i} = \begin{cases} 1 & \delta_{t,i} = \cdots = \delta_{r+2,i} = 0, \delta_{r+1,i} = \delta_{r,i} = 1, \\ & r = 1, \ldots, t-2, \quad t = 2, \ldots, T. \\ 0 & \text{otherwise,} \end{cases}$$

A nonrespondent $y_{t,j}$ with last respondent $y_{r,j}$ is imputed by

$$\tilde{y}_{t,j} = \hat{\phi}_{t,r}(y_{r,j}).$$

The Markov chain assumption (2) ensures that using previously imputed values $\tilde{y}_{t,i}$ as outcomes in (8) produces an asymptotically valid estimator of $\phi_{t,r}$ (see result (11) in the Appendix).

3.3 An illustration

To illustrate the previous described imputation process and how nonresponse patterns are grouped into imputation cells, we consider imputation at time point $t = 4$ (Table 1). The horizontal direction in Table 1 corresponds to 4 time points and the vertical direction corresponds to different nonresponse patterns, where each pattern is represented by a 4-dimensional vector of 0's and 1's with 0 indicating a nonrespondent and 1 indicating a respondent. There are a total of $2^{T-1} = 2^3 = 8$ nonresponse patterns. According to the previously described imputation process, at step 1, we consider nonrespondents at time 4 with last respondents at time 3, which are patterns 3 and 4. According to imputation model (5), we fit a regression using data in patterns 7 and 8 indicated by + (used as predictors) and × (used as outcomes) in the block in Table 1 under title step 1. Then, imputed values (indicated by ○) are obtained from the fitted regression using data indicated by * as predictors in the block under title step 1. Next, we focus on the block in Table 1 under title step 2. The nonrespondents at $t = 4$ with last respondents at time 2 are those in pattern 2. According to imputation model (7), we fit a regression using data in pattern 3 indicated by + (used as predictors) and ⊗ (previously imputed values used as outcomes). Then, imputed values (indicated by ○) are obtained from the fitted regression using data indicated by * as predictors. Finally, we focus on the block in Table 1 under title step 3. The nonrespondents at $t = 4$ with last respondents at time 1 are those in pattern 1. According to imputation model (7), we fit a regression using data in pattern 2 indicated by + (used as predictors) and ⊗ (previously imputed values used as outcomes). Then, imputed values (indicated by ○) are obtained from the fitted regression using data indicated by * as predictors.

Table 1
Illustration of imputation process at $t = 4$

	Step 1: $r = 3$				Step 2: $r = 2$				Step 3: $r = 1$			
	Time				Time				Time			
Pattern	1	2	3	4	1	2	3	4	1	2	3	4
(1,0,0,0)									*			○
(1,1,0,0)						*		○	+			⊗
(1,1,1,0)			*	○		+		⊗				
(1,0,1,0)			*	○								
(1,0,0,1)												
(1,1,0,1)												
(1,0,1,1)			+	×								
(1,1,1,1)			+	×								

+: observed data used in regression fitting as predictors
×: observed data used in regression fitting as responses
⊗: imputed data used in regression fitting as responses
*: observed data used as predictors in imputation
○: imputed values

4. Estimation of population means using imputed data

Let $\bar{Y}_t$ be the finite population mean at time point *t*. The sample mean based on observed and imputed data is

$$\hat{\bar{Y}}_t = \sum_{i \in S} w_i \tilde{y}_{t,i}, \tag{9}$$

where $\tilde{y}_{t,i}$ is equal to the observed value if $\delta_{t,i} = 1$ and is an imputed value if $\delta_{t,i} = 0$. We now establish that, as an estimator for the population mean at time point t, $\hat{\bar{Y}}_t$ in (9) is consistent and asymptotically normal under the asymptotic frame work described in Section 2.

Theorem 1. Assume (1)-(2) and (4), and the asymptotic frame work described in Section 2. Assume further the following regularity conditions:

(C1) $E(y_t^2) < \infty, t = 1, \ldots, T$.

(C2) $0 < P(I_{t,r} = 1) < 1$ and $E[\sigma_{t,r}^2(y_r)/p_{t,r}(y_r)] < \infty$, where $p_{t,t-1}(x) = P(\delta_t = 1 \mid y_{t-1} = x, \delta_{t-1} = 1)$, $p_{t,r}(x) = P(\delta_t = 0 \mid y_r = x, \delta_r = \delta_{r+1} = 1, \delta_{t-1} = \cdots = \delta_{r+2} = 0)$, $r = 1, \ldots, t-2$, $\sigma_{t,r}^2(x) = \mathrm{Var}(y_t \mid y_r = x, I_{t,r} = 1)$, $I_{t,r}$ is the same as $I_{t,r,i}$ with $\delta_{t,i}$'s replaced by δ_t's, $r = 1, \ldots, t-1, t = 2, \ldots, T$.

(C3) $\phi_{t,r}(x)$ and $g_{t,r}(x) = p_{t,r}(x) f_r(x)$ have bounded second derivatives such that

$$E[\{\sigma_{t,r}^2(y_r) + \phi_{t,r}(y_r)\} g''_{t,r}(y_r) \times \{1 - p_{t,r}(y_r)\} / \sqrt{g_{t,r}(y_r)} < \infty$$

and

$$E[\{\phi_{t,r}(y_r) g''_{t,r}(y_r)^2 + \phi'_{t,r}(y_r) g'_{t,r}(y_r)\} \times \{1 - p_{t,r}(y_r)\} / g_{t,r}(y_r)] < \infty,$$

where $f_r(x)$ is the probability density function of y_r, $r = 1, \ldots, t-1, t = 2, \ldots, T$.

(K) The kernel function κ is a bounded and symmetric probability density function on the real line with finite second moment.

(B) The bandwidth h satisfies $nh^2/(\log n)^2 \to \infty$ and $nh^4 \to 0$ as $n \to \infty$.

Then, for $t = 1, \ldots, T$,

$$\sqrt{n}(\hat{\bar{Y}}_t - \mu_t) \to_d N(0, \sigma_t^2), \tag{10}$$

where $\mu_t = E(y_t)$, σ_t^2 is an unknown parameter, and $\to_d$ denotes convergence in distribution with respect to the joint distribution of $(y_{1,i}, \ldots, y_{T,i}, \delta_{2,i}, \ldots, \delta_{T,i})$ and sampling (model and design).

The proof is given in the Appendix. Conditions (K) and (B) are exactly the same as those in Cheng (1994) and conditions (C1)-(C3) are the same as those in Cheng (1994) applied to units defined by the right hand sides of (5)-(7).

Because of the complexity of the imputation, it is difficult to obtain an explicit form of σ_t^2 in (9). We consider the bootstrap method. A correct bootstrap can be obtained by applying the imputation process in each of the bootstrap samples, *i.e.*, by imputing each bootstrap data set exactly the way the original data set is imputed (Shao and Sitter 1996). More specifically, we proceed as follows.

1. Within each imputation class, draw a bootstrap sample as a simple random sample with replacement from the sample, where the bootstrap sample size is the same as the number of sampled units in the imputation class. Combine the bootstrap samples to form S^*. The bootstrap data set contains all observed data, weights, and response indicators of units in S^*.
2. Apply the proposed imputation procedure to the bootstrap data set. Calculate the bootstrap analogue $\hat{\bar{Y}}_t^*$.
3. Independently repeat the previous steps B times to obtain $\hat{\bar{Y}}_t^{*1}, \ldots, \hat{\bar{Y}}_t^{*B}$. The sample variance of $\hat{\bar{Y}}_t^{*1}, \ldots, \hat{\bar{Y}}_t^{*B}$ is our bootstrap variance estimator for $\hat{\bar{Y}}_t$.

Note that the bootstrap method requires a large amount of repeated computation, which is the price paid for replacing a theoretical derivation of asymptotic variances. One may also use other valid bootstrap methods for survey data described in Shao and Tu (1995, Chapter 6).

Performance of the proposed bootstrap variance estimator is evaluated by simulation in the next section.

5. Simulation

A simulation study was conducted to evaluate the performance of the proposed imputation method in terms of the estimation of the mean of y_t. We considered a sample of size 1,000. Each sample unit has longitudinal data of size $T = 4.$ The population mean values for the 4 time points are 1.33, 1.94, 2.73, and 3.67, respectively. Longitudinal data were generated according to two models. In the first model, $(y_1, \ldots, y_T)$ is multivariate normal and follows the AR(1) model with correlation coefficient 0.9 and standard error 1. In the second model, $(\log y_1, \ldots, \log y_T)$ is multivariate normal and follows the AR(1) model with correlation coefficient 0.9 and standard error 1. All data at $t = 1$ are observed. For $t = 2, \ldots, T,$ nonrespondents were generated using

$$P(\delta_t = 0 \mid y_{t-1}) = \frac{\exp\{1 - 1.2 y_{t-1}\}}{1 + \exp\{1 - 1.2 y_{t-1}\}}$$

for the case of normal data and

$$P(\delta_t = 0 \mid y_{t-1}) = \frac{\exp\{2 - 0.7 y_{t-1}\}}{1 + \exp\{2 - 0.7 y_{t-1}\}}$$

for the case of log-normal data. These nonresponse models were chosen so that the unconditional probabilities of non-response are about the same in the two cases (see Table 2).

For comparison, we included five estimators in the simulation: the sample mean of complete data, which is used as the gold standard; the sample mean of respondents that ignores nonrespondents; the sample mean based on simple linear regression imputation described in Section 1; the sample mean based on censoring (as described in Section 1) and linear regression imputation for monotone missing data (Paik 1997); and the sample mean based on our proposed imputation procedure. In the nonparametric regression described in Sections 3.1 and 3.2, the standard normal density was used as the kernel $\kappa(x)$ and the bandwidth h was chosen to be around $4n^{-2/5}$, which was used in the simulation in Cheng (1994).

Table 2
Unconditional probabilities of nonresponse in the simulation study

	Nonresponse pattern		Nonresponse probability			
			The normal case		The log-normal case	
$t = 3$	Monotone	(1,0,0)	0.14		0.16	
		(1,1,0)	0.12	0.26	0.09	0.25
	Intermittent	(1,0,1)	0.25	0.25	0.20	0.20
	Complete	(1,1,1)	0.49	0.49	0.55	0.55
$t = 4$	Monotone	(1,0,0,0)	0.04		0.06	
		(1,1,0,0)	0.02		0.02	
		(1,1,1,0)	0.04	0.10	0.02	0.10
	Intermittent	(1,0,0,1)	0.10		0.10	
		(1,0,1,0)	0.04		0.03	
		(1,0,1,1)	0.21		0.17	
		(1,1,0,1)	0.10	0.45	0.06	0.36
	Complete	(1,1,1,1)	0.45	0.45	0.54	0.54

Tables 3-4 report (based on 1,000 simulation runs) the relative bias and variance of mean estimators, the bootstrap variance estimator (based on 200 bootstrap replications), the coverage probability of approximate 95% confidence intervals (CI) obtained using point estimator $\pm 1.96 \times \sqrt{\text{bootstrap variance}}$, and the length of CI. The results in Tables 3-4 can be summarized as follows.

1. Bias. The proposed imputation method produces estimators with negligible bias in all cases under consideration. The sample mean of respondents only is clearly biased unless $t = 1$. Although in some cases the values of the bias are small, the bias leads to very low coverage probability of the CI, because the variance of the sample mean is much smaller than its squared bias. The simple linear regression imputation method is correct only when $t = 2$ and data are normally distributed. Its relative bias at $t = 3$ in the normal case is very small, but at $t = 4$, it has a relative bias of 1.1% that leads to a coverage probability 76.3% only for its CI. The method of censoring and linear regression imputation is correct in the normal case and has little bias. In the log-normal case, however, both the simple linear regression imputation and the method of censoring with linear regression imputation are biased, due to the fact that the regression functions are not linear.
2. The bootstrap and CI. The bootstrap variance as an estimator of the variance of the mean estimator performs well in all cases, even when the mean estimator is biased. The related CI has a coverage probability close to the nominal level 95% when the mean estimator has no bias.
3. Proposed imputation versus censoring. When censoring and linear regression imputation is used, the mean estimator is biased in the log-normal case and, thus, the proposed imputation method is clearly better. In the normal case, both methods are correct. However, the results in Table 3 show the effect of discarding observed data. When $t = 2$, censoring is better than the proposed imputation method, because no unit is actually censored and the censoring method uses the correct linear regression whereas the proposed imputation method fits a nonparametric regression. When $t = 3$, censoring is about the same as the proposed imputation method but when $t = 4$, censoring is a lot worse than the proposed imputation method. From Table 2, on average 25% sample units are censored when $t = 3$ and 45% sample units are censored when $t = 4$. The gain in using a correct linear regression is not enough to compensate the effect of discarding observed data, especially when $T = 4$.

Table 3
Simulation results for estimation of means (Normal case)

Method	Quantity	$t = 1$	$t = 2$	$t = 3$	$t = 4$
complete data	relative bias	0.0%	0.0%	0.0%	0.0%
	variance $\times 10^3$	0.962	0.981	1.052	1.033
	bootstrap variance	1.002	1.002	1.002	1.006
	coverage prob of CI	95.4%	94.9%	94.5%	94.5%
	length of CI	0.124	0.124	0.124	0.124
respondents	relative bias		16.8%	8.3%	3.5%
	variance $\times 10^3$		1.319	1.240	1.051
	bootstrap variance		1.364	1.178	1.062
	coverage prob of CI		0.0%	0.0%	2.4%
	length of CI		0.145	0.134	0.128
simple linear regression imputation	relative bias		0.0%	0.0%	1.1%
	variance $\times 10^3$		1.121	1.434	1.185
	bootstrap variance		1.172	1.466	1.192
	coverage prob of CI		94.9%	94.7%	76.3%
	length of CI		0.134	0.150	0.135
censoring and linear regression imputation	relative bias		0.0%	0.0%	0.0%
	variance $\times 10^3$		1.121	1.437	1.642
	bootstrap variance		1.172	1.476	1.819
	coverage prob of CI		94.9%	94.7%	96.1%
	length of CI		0.134	0.150	0.167
proposed imputation	relative bias		0.2%	0.3%	0.2%
	variance $\times 10^3$		1.196	1.438	1.264
	bootstrap variance		1.231	1.401	1.224
	coverage prob of CI		95.0%	93.7%	94.1%
	length of CI		0.137	0.146	0.137

Table 4
Simulation results for estimation of means (Log-normal case)

Method	Quantity	$t=1$	$t=2$	$t=3$	$t=4$
complete data	relative bias	0.0%	0.0%	0.0%	0.0%
	variance	0.069	0.172	0.383	1.074
	bootstrap variance	0.067	0.161	0.418	1.138
	coverage prob of CI	94.4%	93.8%	94.9%	94.6%
	length of CI	1.008	1.557	2.511	4.142
respondents	relative bias		28.1%	18.8%	10.8%
	variance		0.366	0.614	1.378
	bootstrap variance		0.344	0.668	1.461
	coverage prob of CI		0.1%	2.1%	31.6%
	length of CI		2.267	3.171	4.690
simple linear regression imputation	relative bias		7.0%	12.6%	12.5%
	variance		0.266	0.877	1.589
	bootstrap variance		0.240	0.807	1.611
	coverage prob of CI		71.6%	39.3%	23.2%
	length of CI		1.894	3.481	4.938
censoring and linear regression imputation	relative bias		7.0%	12.1%	13.8%
	variance		0.266	0.874	2.735
	bootstrap variance		0.240	0.836	2.617
	coverage prob of CI		71.6%	43.9%	36.4%
	length of CI		1.894	3.540	6.277
proposed imputation	relative bias		0.1%	0.1%	0.1%
	variance		0.189	0.447	1.119
	bootstrap variance		0.179	0.482	1.236
	coverage prob of CI		94.5%	95.7%	95.6%
	length of CI		1.644	2.697	4.317

6. An example

In the CES introduced in Section 1, data for employment are collected from nonagricultural establishments on a monthly basis. In any particular month after the baseline, the nonresponse rate is about 20-40% and nonresponse is nonmonotone. In CES, it is typically assumed that (1)-(2) hold. In fact, assumption (3) that is stronger than assumption (2) is often assumed (Butani, Harter and Wolter 1997). We consider a stratified simple random sample from a CES dataset (a subset of a sample from the 1980's). Stratum sizes, sample size by stratum, and nonresponse rate by stratum are listed in Table 5. For each imputation method, imputation is carried out within a group of strata (group 1 = strata 1-4; group 2 = strata 5-7; group 3 = strata 8-11; group 4 = stratum 12; group 5 = strata 13-15; group 6 = stratum 16).

To estimate the employment counts from month 1 (baseline) to month 8, we applied the five methods in the simulation study in Section 5. The kernel and bandwidth in nonparametric regression were the same as those in the simulation (Section 5). Since population employment counts are obtained once a year from Unemployment Insurance administrative records, nonrespondents in any month actually become available later so that the sample mean of complete data is available as a standard. The sample means based on different methods are reported in Table 6 together with their bootstrap variance estimates (based on bootstrap sample size 200). For the sample means based on respondents and three imputation methods, we also computed the estimated relative bias defined as (sample mean/sample mean of complete data) -1.

The result in Table 6 shows that the sample mean based on the proposed imputation method is very comparable to the sample mean from the complete data, whereas the sample mean of respondents is clearly biased. Due to the fact that nonresponse is nonignorable, the simple linear regression imputation shows some bias starting from month 4, although the estimated relative bias is at most 5.5% in absolute value. The method of censoring with linear regression imputation has some bias after month 4, probably due to the fact that data are not normally distributed so that fitting linear regression is not correct. Furthermore, it has larger estimated variances compared with the proposed method, indicating the effect of discarding observed data.

Table 5
Stratum size, sample size, and nonresponse rate in the CES example

Stratum	Stratum size	Sample size	Nonresponse percentage $t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
1	223	102	0	32.4	39.2	34.3	27.5	30.4	28.4	33.3
2	1,649	110	0	28.2	31.8	30.0	34.5	33.6	26.4	29.1
3	1,900	120	0	37.5	39.2	34.2	44.2	41.7	40.0	40.8
4	419	98	0	41.8	48.0	35.7	38.8	44.9	38.8	38.8
5	1,947	132	0	37.1	33.3	25.8	25.0	27.3	23.5	28.8
6	2,391	180	0	41.1	36.1	42.8	37.8	39.4	39.4	38.3
7	5,365	256	0	35.2	34.0	33.6	36.7	35.2	40.6	39.1
8	2,330	201	0	30.3	36.8	40.3	34.8	37.3	37.3	37.8
9	1,164	113	0	35.4	29.2	33.6	30.1	29.2	32.7	33.6
10	593	106	0	37.7	44.3	40.6	47.2	41.5	37.7	32.1
11	2,222	182	0	24.2	26.4	27.5	27.5	28.0	20.3	27.5
12	6,880	512	0	40.0	39.6	40.6	41.0	41.4	39.8	38.9
13	2,373	160	0	36.9	40.6	36.2	33.8	39.4	30.0	36.9
14	50	42	0	40.5	38.1	28.6	45.2	33.3	33.3	31.0
15	4,100	241	0	36.5	38.6	34.4	34.9	42.7	33.2	32.8
16	3,951	412	0	37.9	36.9	36.9	38.1	40.3	40.5	39.3

Table 6
Estimates in the CES example

Method	Quantity	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
complete data	mean	38.05	38.41	38.70	38.95	39.16	38.91	38.79	38.81
	$\widehat{\text{var}}$	0.805	0.814	0.828	0.830	0.990	0.832	0.822	0.852
respondents	mean		54.03	54.31	54.15	55.08	55.15	54.20	54.50
	$\widehat{\text{bias}}$		40.7%	40.3%	39.0%	40.7%	41.7%	39.7%	40.4%
	$\widehat{\text{var}}$		1.647	1.488	1.506	1.990	1.708	1.413	1.491
simple linear	mean		38.45	38.72	39.33	38.81	41.04	40.54	39.79
regression imputation	$\widehat{\text{bias}}$		0.1%	0.1%	1.0%	-0.9%	5.5%	4.5%	2.5%
	$\widehat{\text{var}}$		0.821	0.834	0.866	0.963	1.979	1.465	1.008
censoring and linear	mean		38.45	38.71	39.17	38.25	40.46	40.34	40.30
regression imputation	$\widehat{\text{bias}}$		0.1%	0.0%	0.6%	-2.3%	4.0%	4.0%	3.8%
	$\widehat{\text{var}}$		0.821	0.833	0.881	1.289	1.497	1.630	1.660
proposed imputation	mean		38.37	38.68	38.97	39.10	39.05	38.72	38.88
	$\widehat{\text{bias}}$		-0.1%	-0.1%	0.0%	-0.1%	0.4%	-0.2%	0.2%
	$\widehat{\text{var}}$		0.813	0.834	0.837	1.019	0.962	0.924	0.910

$\widehat{\text{bias}}$: (sample mean / sample mean of complete data) - 1
$\widehat{\text{var}}$: bootstrap variance estimate

7. Concluding remarks

For longitudinal data with nonmonotone nonresponse, we propose an imputation method under the assumptions that the nonresponse mechanism is last-value-dependent and the longitudinal data follow a Markov chain. Our method is nonparametric and produces consistent and asymptotically normally distributed estimators of population means. Because the asymptotic variances of the estimators of population means are very complicated, we propose a simple bootstrap method for variance estimation. Some simulation results show that the proposed method works well. The CES is our motiving example and is used for illustration.

In general, nonresponse of data at time point t may depend not only on the last value at time $t-1$, but also on other past values at time points prior to $t-1$. Furthermore, the longitudinal data may not be a Markov chain, *i.e.*, there may be long time dependence among data from different time points. In either case, our proposed method is not applicable. A general method is still under development.

Appendix

Proof of (5)

Let $L(\xi)$ denote the distribution of ξ and $L(\xi \mid \zeta)$ denote the conditional distribution of ξ given ζ. Then,

$$L(y_t \mid y_{t-1}, \delta_t = 0, \delta_{t-1} = 1)$$

$$= \frac{L(y_t, y_{t-1}, \delta_t = 0, \delta_{t-1} = 1)}{L(y_{t-1}, \delta_t = 0, \delta_{t-1} = 1)}$$

$$= \frac{L(\delta_t = 0 \mid y_t, y_{t-1}, \delta_{t-1} = 1) L(y_t, y_{t-1}, \delta_{t-1} = 1)}{L(\delta_t = 0 \mid y_{t-1}, \delta_{t-1} = 1) L(y_{t-1}, \delta_{t-1} = 1)}$$

$$= L(y_t \mid y_{t-1}, \delta_{t-1} = 1),$$

where the third equality follows from (1). Similarly, we can show that

$$L(y_t \mid y_{t-1}, \delta_t = 1, \delta_{t-1} = 1) = L(y_t \mid y_{t-1}, \delta_{t-1} = 1).$$

Hence, $L(y_t \mid y_{t-1}, \delta_t = 0, \delta_{t-1} = 1) = L(y_t \mid y_{t-1}, \delta_t = 1, \delta_{t-1} = 1)$ and result (5) follows.

Proof of (7)

Using the same notation as in the previous proof, we have

$$\begin{aligned}
&L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+1} = 0, \delta_r = 1)\\
&\quad= \frac{L(y_t, y_r, \delta_t = \cdots = \delta_{r+1} = 0, \delta_r = 1)}{L(y_r, \delta_t = \cdots = \delta_{r+1} = 0, \delta_r = 1)}\\
&\quad= \frac{L(\delta_{r+1} = 0 \mid y_t, y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1)}{L(\delta_{r+1} = 0 \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1)}\\
&\qquad\times \frac{L(y_t, y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1)}{L(y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1)}\\
&\quad= L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1),
\end{aligned}$$

where the last equality follows from (1). Similarly, we can show that

$$\begin{aligned}
&L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1)\\
&\qquad= L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_r = 1).
\end{aligned}$$

Hence,

$$\begin{aligned}
&L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+1} = 0, \delta_r = 1)\\
&\qquad= L(y_t \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1)
\end{aligned}$$

and result (7) follows.

Proof of theorem 1

Let $t(= 2, \ldots, T)$ be fixed and $n_{t,r} =$ the number of units with $I_{t,r,i} = 1, r = 1, \ldots, t-1$. We first show that, for $r < t-1$,

$$E(\phi_{t,r+1}(y_{r+1}) \mid y_r, I_{t,r} = 1) = E(y_t \mid y_r, I_{t,r} = 1). \quad (11)$$

Under assumption (1),

$$\begin{aligned}
&L(y_t \mid y_{t-1}, \ldots, y_1, \delta_{t-1}, \ldots, \delta_1)\\
&= \frac{L(\delta_{t-1} \mid y_t, y_{t-1}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1)}{L(\delta_{t-1} \mid y_{t-1}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1)} L(y_t \mid y_{t-1}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1)\\
&= L(y_t \mid y_{t-1}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1)\\
&\ldots\\
&= L(y_t \mid y_{t-1}, \ldots, y_1)\\
&= L(y_t \mid y_{t-1}),
\end{aligned}$$

where the last equality follows from assumption (2). Then

$$\begin{aligned}
&L(y_t \mid y_{t-1}, \delta_{t-1})\\
&\quad= \int L(y_t \mid y_{t-1}, \ldots, y_1, \delta_{t-1}, \ldots, \delta_1)\\
&\qquad dL(y_{t-2}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1 \mid y_{t-1}, \delta_{t-1})\\
&\quad= L(y_t \mid y_{t-1}) \int dL(y_{t-2}, \ldots, y_1, \delta_{t-2}, \ldots, \delta_1 \mid y_{t-1}, \delta_{t-1})\\
&\quad= L(y_t \mid y_{t-1})
\end{aligned}$$

and

$$\begin{aligned}
&L(y_t, \delta_t \mid y_{t-1}, \ldots, y_1, \delta_{t-1}, \ldots, \delta_1)\\
&\qquad= L(\delta_t \mid y_t, y_{t-1}, \ldots, y_1, \delta_{t-1}, \ldots, \delta_1)\\
&\qquad\quad\times L(y_t \mid y_{t-1}, \ldots, y_1, \delta_{t-1}, \ldots, \delta_1)\\
&\qquad= L(\delta_t \mid y_{t-1}) L(y_t \mid y_{t-1})\\
&\qquad= L(\delta_t \mid y_t, y_{t-1}, \delta_{t-1}) L(y_t \mid y_{t-1}, \delta_{t-1})\\
&\qquad= L(y_t, \delta_t \mid y_{t-1}, \delta_{t-1}).
\end{aligned}$$

Hence, $\{(y_t, \delta_t), t = 1, \ldots, T\}$ is a Markov chain. Consequently,

$$\begin{aligned}
&E(y_t \mid y_{r+1}, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = 1)\\
&\qquad= E(y_t \mid y_{r+1}, y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1).
\end{aligned}$$

Then, the left hand side of (11) is equal to

$$\begin{aligned}
&E[E(y_t \mid y_{r+1}, y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1) \mid\\
&\qquad\qquad y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1]\\
&\qquad= E(y_t \mid y_r, \delta_t = \cdots = \delta_{r+2} = 0, \delta_{r+1} = \delta_r = 1),
\end{aligned}$$

which is the right hand side of (11).

It follows from the construction of $\tilde{y}_{t,i}$ described in Section 3, result (11), and the proof of Theorem 2.1 in Cheng (1994) that

$$\sqrt{n_{t,r}}(\bar{y}_{t,r} - \mu_{t,r}) \to_d N(0, \sigma^2_{t,r})$$

for some $\sigma^2_{t,r}$, where $\bar{y}_{t,r} = 1/n_{t,r} \sum_{i=1}^{n} I_{t,r,i}\, \tilde{y}_{t,i}$ and $\mu_{t,r} = E[\phi_{t,r}(y_r) \mid I_{t,r} = 1]$, $r = 1, \ldots, t-1$. The result follows from

$$\begin{aligned}
\mu_t &= \sum_{r=1}^{t-1} E(y_t \mid I_{t,r} = 1) P(I_{t,r} = 1)\\
&= \sum_{r=1}^{t-1} E[E(y_t \mid y_r, I_{t,r} = 1) \mid I_{t,r} = 1] P(I_{t,r} = 1)\\
&= \sum_{r=1}^{t-1} E[\phi_{t,r}(y_r) \mid I_{t,r} = 1] P(I_{t,r} = 1)\\
&= \sum_{r=1}^{t-1} \mu_{t,r} P(I_{t,r} = 1)
\end{aligned}$$

and

$$\sqrt{n}(\hat{\hat{Y}}_t - \mu_t) = \sum_{r=1}^{t-1}\left[\frac{\sqrt{n_{t,r}}}{\sqrt{n}}\sqrt{n_{t,r}}\,(\bar{y}_{t,r} - \mu_{t,r}) + \mu_{t,r}\sqrt{n}\left(P(I_{t,r}=1) - \frac{n_{t,r}}{n}\right)\right].$$

Acknowledgements

The authors would like to thank two referees and one associate editor for their helpful comments and suggestions. The research was partially supported by the NSF Grants DMS-0404535 and SES-0705033.

References

Bickel, P.J., and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, 470-482.

Butani, S., Harter, R. and Wolter, K. (1997). Estimation procedures for the Bureau of Labor Statistics Current Employment Statistics Program. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 523-528.

Cheng, P.E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 81-87.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing data. *Survey Methodology*, 12, 1-16.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 1010-1019.

Little, R.J., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc.

Paik, M.C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association*, 1320-1329.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995). Analysis of semi-parametric regression method for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106-121.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. New York: CRC Press.

Shao, J., and Sitter, R.R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 1278-1288.

Shao, J., and Tu, D. (1995). The Jackknife and Bootstrap. New York: Springer.

Troxel, A.B., Harrington, D.P. and Lipsitz, S.R. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics*, 47, 425-438.

Troxel, A.B., Lipsitz, S.R. and Brennam, T.A. (1997). Weighted estimating equations with nonignorable missing response data. *Biometrics*, 53, 857-869.

Troxel, A.B., Lipsitz, S.R. and Harrigton, D.P. (1998). Marginal models for the analysis of longitudinal measurements with non-ignorable non-monotome missing data. *Biometrika*, 85, 661-672.

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, 89-96.

Xu, J. (2007). Methods for intermittent missing responses in longitudinal data. Ph.D. Thesis, Department of Statistics, University of Wisconsin-Madison.