Statistica Sinica **35** (2025), 131-150
doi:https://doi.org/10.5705/ss.202022.0214

SEMIPARAMETRIC ESTIMATION OF NON-IGNORABLE MISSINGNESS WITH REFRESHMENT SAMPLE

Jianfei Zheng, Jing Wang, Lan Xue* and Annie Qu

BeiGene, University of Illinois at Chicago, Oregon State University and University of California, Irvine

Abstract: The problem of missing data is common in longitudinal data analysis and poses methodological challenges in terms of providing unbiased estimation and statistical inference, owing to informative missingness. In such cases, it is crucial to correctly identify and appropriately incorporate the missing mechanism into estimation and inference procedures. Traditional methods, such as the complete-case analysis and imputation methods, are designed to deal with missing data under unverifiable assumptions of missing completely at random and missing at random. We focus on identifying and estimating missing parameters under the non-ignorable missing assumption, using refreshment samples from two-wave panel data. Specifically, we propose a full-likelihood approach when a parametric model is specified for the joint distribution of two-wave data. When such a model is unavailable, we propose a semiparametric method to estimate the attrition parameters, with marginal density estimates obtained using an additional refreshment sample. We derive several asymptotic properties of the semiparametric estimators, and demonstrate their numerical performance using simulations. We further propose an inference on bootstrapping, and assess it using simulations. Lastly, a real-data application is provided based on the Netherlands Mobility Panel study.

Key words and phrases: Additive non-ignorable missing, asymptotic normality, kernel density estimator, Netherlands Mobility Panel, wave data.

1. Introduction

Panel or longitudinal studies are widely used in scientific fields to assess changes at both population and individual levels. However, longitudinal studies often suffer from attrition, where some subjects are unable to respond to follow-up studies, resulting in incomplete panel data and significant challenges for traditional statistical methods. For example, the Netherlands Institute for Transport Policy Analysis (Hoogendoorn-Lanser, Schaap and OldeKalter, 2015) has been conducting the Netherlands Mobility Panel (NMP) since 2013. The panel currently involves two waves of data collection, with the initial wave consisting of 2,380 households. For the second wave, only 1,685 households remained after almost 30% dropped out. Bias can be introduced in statistical inference if attrition is ignored and the missingness is systematically related to

*Corresponding author. E-mail: xuel@science.oregonstate.edu

the responses. Therefore, understanding the missing mechanism is crucial when making statistical inferences about populations.

Different models have been proposed to explain missingness (Rubin, 2004), such as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR assumes the missingness is independent of all the variables in the data, both observed and missing, whereas MAR allows the missing mechanism to depend on variables that are always observed. MNAR further relaxes the assumption for the missing mechanism, and assumes the missingness depends on both observed and unobserved variables. Numerous statistical methods have been developed to allow valid estimations and inferences under these missingness assumptions.

Unfortunately, partially observed panel data alone cannot distinguish among the various missing mechanisms, and the aforementioned missingness assumptions are often unverifiable. A violation of the assumptions could lead to biased estimation and inference (Deng et al., 2013), and the MNAR model has identification issues, because the panel data alone are often not sufficient to make inferences about populations (Rubin, 1976, 2004; Hirano et al., 2001; Fitzmaurice et al., 2008). Miao, Ding and Geng (2016) provide sufficient conditions for model identifiability when the response follows a normal or a normal mixture distribution. Furthermore, d'Haultfoeuille (2010), under a completeness assumption, and Wang, Shao and Kim (2014), using the generalized method of moments, establish sufficient identifiability conditions for general data-generating processes by introducing an instrumental variable. Assuming a semiparametric model on the response mechanism, based on estimating equations, Morikawa and Kim (2021) provide a sufficient condition for its identifiability without needing the instrumental variable assumption.

Hirano et al. (2001) were the first to explore using refreshment samples to improve the estimation and inference of the attrition process. A refreshment sample is a common sampling strategy of collecting a new random sample from the target population during follow-up waves when attrition occurs. Many large panel studies now routinely include refreshment samples (Deng et al., 2013). For instance, many longitudinal studies of the National Center for Education Statistics, including the Early Childhood Longitudinal Study (Asigbee, Whitney and Peterson, 2018) and the National Educational Longitudinal Study (Ingels et al., 2014), refill their samples once or multiple times during a study. The NMP completed its initial data survey in 2013, after which a follow-up survey was administered in 2014 that included a refreshment sample.

Refreshment samples provide an inexpensive way to improve the quality of longitudinal data, and various methods have been developed to estimate the attrition process using a refreshment sample. Hirano et al. (2001) propose an additive non-ignorable model that takes MCAR and MAR models as special cases to gain insights and make inferences for the attrition process. They provide the fundamental identification theory and develop an estimation procedure for a two-wave binary response. Nevo (2003) uses a refreshment sample to compute sampling weights so that the moments of the weighted data match those observed in the refreshment sample. Bhattacharya (2008) converts Hirano's fundamental identification theory into conditional moment restrictions. A set of nonparametric regressions with B-splines are used to construct the objective function for the parameter estimation. Deng (2012) and Deng et al. (2013) extended the additive non-ignorable model by including two refreshment samples to handle three-wave binary response data, using a fully Bayesian approach and a Markov chain Monte Carlo estimation. Similarly, Si, Reiter and Hillygus (2015) present a semiparametric additive non-ignorable model for analyzing multivariate categorical responses in a two-wave panel with one refreshment sample. They use the additive non-ignorable model for the attrition process and model the multinomial survey responses using a Dirichlet process mixture.

This paper proposes two new approaches for handling MNAR data in a two-wave panel with one refreshment sample. The first method is a fully parametric method based on likelihood. Inferences for the population use maximum likelihood estimators, and we use an adaptive Gaussian quadrature to overcome the integration difficulty introduced by the missing data in the construction of the likelihood. The second method is a semiparametric approach that uses the kernel density estimator as the nonparametric component, and the additive non-ignorable attrition model (Hirano et al., 2001) as the parametric component. The proposed semiparametric method is based on matching the marginal densities recovered from the panel data with the observed marginal densities from the first wave and the refreshment sample. The proposed method is easy to implement and fast to compute. When the likelihood is specified correctly, the full-likelihood approach provides the most efficient estimators and acts as a benchmark for MNAR data analysis methods in a two-wave panel. However, when the likelihood is misspecified, the full-likelihood method results in bias and invalid inferences. On the other hand, the semiparametric method is more robust and flexible in terms of the distributional specification and provides consistent inferences for the attrition process under different population conditions. Simulation results support the finding that the kernel density-based semiparametric estimators exhibit better numerical performance than that of the method proposed by Bhattacharya (2008).

The first contribution of this study follows from combining the advantages of Hirano's fundamental identification theory (Hirano et al., 2001) with kernel density estimators. The proposed semiparametric method does not require a specification of the joint distribution of the data and provides a unified estimation procedure for the additive MNAR model. The second contribution is the theoretical justification of the proposed estimators. While no asymptotic justification is given in Hirano et al. (2001) or Deng et al. (2013), we show that the semiparametric estimator is consistent and asymptotically normal, and develop inference tools are developed for testing the MCAR and MAR assumptions based on asymptotic formulae and bootstrapping methods. The proposed methods differ fundamentally from those designed for binary data (Hirano et al., 2001; Deng et al., 2013), because the distribution of binary data can be characterized using a few parameters, and the estimation procedure involves only moments. In contrast, the continuous case requires parameters of infinite dimension, creating challenges in both computation and theory development.

The rest of the paper is organized as follows. Section 2 introduces the refreshment sample and the additive non-ignorable model. Section 3 presents methods for the estimation and inference of the attribution parameters. Extensive simulation results are given in Section 4. An application using the NMP is discussed in Section 5. Finally, Section 6 concludes the paper.

2. Refreshment Sample and Models

In the presence of missingness, it is often assumed that the data are missing completely at random (MCAR) or missing at random (MAR). However, these assumptions are untestable given the panel data alone. When the data are MNAR, the missing mechanism often cannot be identified without additional data or information. Hirano et al. (2001) propose using a refreshment sample to resolve this identification problem and to provide an approach for testing the MCAR and MAR assumptions.

A refreshment sample is an additional independent random sample drawn from the population during follow-up waves when attrition starts to occur. Suppose $\{\mathbf{Y}_i = (Y_{i1}, Y_{i2})\}_{i=1}^N$ are independent and identically distributed (i.i.d.) bivariate responses observed on N subjects from a given population. We assume that the responses in the first wave $\{Y_{i1}\}_{i=1}^N$ are fully observed, and that responses in the second wave $\{Y_{i2}\}_{i=1}^N$ are potentially missing. Let W_i be the missingness indicator, with $W_i = 1$ if Y_{i2} is observed, and $W_i = 0$ otherwise. In addition to the panel data, a refreshment sample of size n is observed at the second wave, and is denoted as $\{Y_{i2}^r\}_{i=1}^n$. With the refreshment sample appended to the original data, the data structure is shown in Table 1.

For the two-wave data in Table 1, Hirano et al. (2001) proposed an additive non-ignorable model for the missing mechanism, of the form

$$P(W = 1 \mid y_1, y_2) = g\left\{\kappa_0 + \kappa_1(y_1) + \kappa_2(y_2)\right\}, \tag{2.1}$$

where g is a monotone function bounded in $[0, 1]$, and κ_0, $\kappa_1(\cdot)$, $\kappa_2(\cdot)$ are constant or arbitrary functions. Model (2.1) includes the MCAR and MAR models as special cases. It leads to the MCAR model if both κ_1 and κ_2 are zero, and to the MAR model if only κ_2 is zero. When κ_2 is nonzero, the data are MNAR. Therefore, the model provides a way of testing for MCAR or MAR

Table 1. Two-wave data with refreshment sample.

	Obs	Y_1	Y_2	W
Complete Set	1	Y_{11}	Y_{12}	W_1=1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_c	$Y_{n_c 1}$	$Y_{n_c 2}$	W_{n_c}=1
Incomplete Set	$n_c + 1$	$Y_{(n_c+1)1}$		W_{n_c+1}=0
	$\vdots$	$\vdots$		$\vdots$
	N	Y_{N1}		W_N=0
Refreshment sample	1		Y_{12}^r	
	$\vdots$		$\vdots$	
	n		Y_{n2}^r	

mechanisms by testing for nonzero κ. This model still includes an untestable assumption that the missingness depends additively on the responses, without any interactions. According to Hirano et al. (2001), this is the weakest assumption that is identifiable and estimable using a refreshment sample.

When both Y_{i1} and Y_{i2} are binary, Hirano et al. (2001) provide two fundamental identification constraints for the attrition parameters and propose estimating these parameters using the method of moments. The authors do not provide an implementation of the additive non-ignorable model continuous responses. We aim to extend the approach of Hirano et al. (2001) to estimate the attrition mechanism for continuous responses using the data observed in Table 1.

We assume non-ignorable missingness and an additive non-ignorable attrition model with the logistic regression form

$$P(W = 1 \mid y_1, y_2) = \frac{\exp(\beta_0 + \beta_1 y_1 + \beta_2 y_2)}{1 + \exp(\beta_0 + \beta_1 y_1 + \beta_2 y_2)}, \tag{2.2}$$

where β_0, β_1, and β_2 are attrition parameters. The logistic model is a popular parametric form for describing a missing mechanism (Rubin, 1976; Hirano et al., 2001; Nevo, 2003; Bhattacharya, 2008; Kim, 2009; Little and Rubin, 2019), as is the probit model. Miao, Ding and Geng (2016) provide sufficient conditions for the probit model to be identified when the response variable follows a normal or normal mixture distribution without using a refreshment sample. Our proposed method can be extended to other parametric attrition models, including the probit model. It can also be extended to a more flexible attrition model with either a nonparametric link function or an additive function of y_1 and y_2 without specifying the functional forms of y_1 and y_2.

3. The Proposed Method

We develop two new methods for handling two-wave MNAR data with continuous responses rather than binary responses. These methods use refreshment samples to estimate the unknown attrition parameters in (2.2). We first describe a likelihood-based fully parametric method in subsection 3.1. Then, in subsection 3.2, we introduce a kernel density-based semiparametric method to estimate the attrition parameters based on Hirano's constraints. The asymptotic theory of the semiparametric estimator is developed in subsection 3.3, and we describe hypothesis testing for the attrition parameters and estimating the corresponding power functions in subsection 3.4.

3.1. Full-likelihood parametric method

We estimate the attrition parameters by maximizing the full likelihood function. The first- and second-wave responses, Y_1 and Y_2, are assumed to be bivariate normal. Let $\theta = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})^T$ and $\beta = (\beta_0, \beta_1, \beta_2)^T$ be the unknown parameters in the bivariate normal and the attrition model, respectively. The three subsets of the data in Table 1 contribute to the likelihood independently. Specifically, in the complete set, responses from both waves are observed, and the likelihood of the complete data is $L_c(\theta, \beta) = \prod_{i=1}^{n_c} f(y_{i1}, y_{i2}, W_i = 1 \mid \theta, \beta) = \prod_{i=1}^{n_c} f(y_{i1}, y_{i2}|\theta) P(W_i = 1 \mid y_{i1}, y_{i2}, \beta)$, where $f(y_1, y_2|\theta)$ is the bivariate normal density function. In the incomplete panel, only the first wave is observed, and its contribution to the likelihood is $L_{ic}(\theta, \beta) = \prod_{i=n_c+1}^{N} f(y_{i1}, W_i = 0 \mid \theta, \beta) = \prod_{i=n_c+1}^{N} \int f(y_{i1}, y_2|\theta) P(W_i = 0 \mid y_{i1}, y_2, \beta) dy_2$. In the refreshment sample, only the second wave is observed, and its contribution to the likelihood is $L_r(\theta) = \prod_{i=1}^{n} f_2(y_{i2}^r|\theta)$. Then, the full likelihood is the product of the above three components as $L(\theta, \beta) = L_c(\theta, \beta) L_{ic}(\theta, \beta) L_r(\theta)$. The maximum likelihood estimates $(\widehat{\theta}_{MLE}, \widehat{\beta}_{MLE})$ can be obtained by maximizing the full likelihood $L(\theta, \beta)$ with respect to all parameters.

Calculating the likelihood of the incomplete set is challenging because it requires integrating a joint density for each incomplete data point, and there is no closed-form solution. To address this, we propose using an adaptive Gaussian–Hermite quadrature (Skrondal and Rabe-Hesketh, 2004; Rabe-Hesketh, Skrondal and Pickles, 2005; Skrondal and Rabe-Hesketh, 2009) for the numerical approximation. The Gaussian–Hermite quadrature is a commonly used technique for generalized linear mixed models (Molenberghs and Verbeke, 2005).

In the parametric approach, the refreshment sample helps to identify the parameters θ and β in the observed likelihood $L(\theta, \beta)$. Miao, Ding and Geng (2016) provide sufficient identifiable conditions for a normal response or normal mixture in a probit model. Without using refreshment samples, the model parameters are, in general, unidentifiable (Hirano et al., 2001). Therefore, the parametric method is infeasible in general non-ignorable missingness scenarios.

The maximum likelihood estimators are most efficient if the underlying population and the attrition models are specified correctly. However, a misspecification of either model can lead to biased estimation and inference. In the next section, we introduce a semiparametric method that does not require a specification of the population distribution and extends Hirano's constraints to the continuous response setting. The parametric method serves as a benchmark to assess the performance of the semiparametric method in simulation studies.

3.2. Kernel density based semiparametric method

Our approach is motivated by the identification equations in Hirano et al. (2001). Let $f(y_1, y_2 \mid W = 1)$ be the joint density of (Y_1, Y_2) on the complete panel, and $f(y_1, y_2)$ be the joint density in the population. When the missing mechanism is specified correctly, we can reconstruct the unobserved joint density $f(y_1, y_2)$ from the observed counterpart $f(y_1, y_2 \mid W = 1)$ by $f(y_1, y_2) = \{P(W = 1)/P(W = 1|y_1, y_2)\}f(y_1, y_2 \mid W = 1)$. As a result, for marginal densities, we have

$$\int \frac{P(W=1)}{P(W=1|y_1,y_2)} f(y_1,y_2 \mid W=1)dy_2 = f_1(y_1),$$
$$\int \frac{P(W=1)}{P(W=1|y_1,y_2)} f(y_1,y_2 \mid W=1)dy_1 = f_2(y_2), \tag{3.1}$$

where f_1 and f_2 are the marginal densities for Y_1 and Y_2 respectively. Our main estimation idea is to find the values of β that correctly transform the joint density in the complete set $f(y_1, y_2 \mid W = 1)$ back into the joint density in the population $f(y_1, y_2)$.

The estimation starts with a two-dimensional kernel density estimator for $f(y_1, y_2 \mid W = 1)$. For any $\mathbf{y} = (y_1, y_2)^T$, the kernel density estimator is $\widehat{f}_H(\mathbf{y} \mid W = 1) = (1/n_c)\sum_{i=1}^{n_c} K_H(\mathbf{y} - \mathbf{Y}_i)$, where $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$, for $i = 1, 2, \ldots, n_c$, are data points in the complete set; H is a 2×2 bandwidth matrix which is symmetric and positive definite; and $K_H(\mathbf{y}) = |H|^{-1/2}K(H^{-1/2}\mathbf{y})$, where K is the bivariate normal kernel function defined as $K(\mathbf{y}) = (2\pi)^{-1}\exp(-\mathbf{y}^T\mathbf{y}/2)$.

In addition, $P(W = 1)$ can be estimated consistently by $\widehat{P}(W = 1) = n_c/N$. For a given $\beta = (\beta_0, \beta_1, \beta_2)^T$, an estimator for the joint density is given as $\widetilde{f}(y_1, y_2 \mid \beta) = \widehat{P}(W = 1)\widehat{f}_H(y_1, y_2 \mid W = 1)/\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)$.

We can compute the marginal densities of Y_1 and Y_2 by numerically integrating the joint distribution $\widetilde{f}(y_1, y_2 \mid \beta)$. In particular, for a given y_1, the marginal density of Y_1 can be computed as $\widetilde{f}_1(y_1 \mid \beta) = \int \widetilde{f}(y_1, y_2 \mid \beta)dy_2$. For a given y_2, $\widetilde{f}_2(y_2 \mid \beta)$ is defined similarly. Due to missingness, we use the refreshment sample rather than the data observed in the second wave to generate the range of Y_2 for the grid points. The resulting marginal density estimates $\widetilde{f}_1(y_1 \mid \beta)$ and $\widetilde{f}_2(y_2 \mid \beta)$ are the semiparametric estimators, which rely on the parametric specification of the attrition model. They are consistent estimates of the true marginal densities only when the attrition model is specified correctly.

The marginal densities on the right-hand side of Equation (3.1) can be estimated directly from the first wave and the refreshment sample. Let $\{Y_{i1}\}_{i=1}^{N}$ be the data from the first wave and $\{Y_{i2}^{r}\}_{i=1}^{n}$ be the refreshment sample. We define one-dimensional kernel density estimators as $\widehat{f}_1(y_1) = \sum_{i=1}^{N} K_{h_1}(y_1 - Y_{i1})/N$, and $\widehat{f}_2(y_2) = \sum_{i=1}^{n} K_{h_2}(y_2 - Y_{i2}^{r})/n$, where K is the univariate density function, and $K_{h_i}(y) = h_i^{-1} K(y/h_i)$, with h_i being the corresponding bandwidth for $i = 1, 2$. In our simulation and numerical studies, we use the plug-in method to select the bandwidths in the kernel density estimators and implement it using the R function *hpi* in the *ks* package.

The estimator $\widehat{\beta}$ of the attrition parameters is defined as the minimizer of the objective function $M_{N,n}(\beta)$ with

$$
\begin{aligned}
M_{N,n}(\beta) &= M_N(\beta) + M_n(\beta) \\
&= \frac{1}{N}\sum_{i=1}^{N} e_{i1}^2 \left\{ \widetilde{f}_1(Y_{i1} \mid \beta) - \widehat{f}_1(Y_{i1}) \right\}^2 + \frac{1}{n}\sum_{i=1}^{n} e_{i2}^2 \left\{ \widetilde{f}_2(Y_{i2}^{r} \mid \beta) - \widehat{f}_2(Y_{i2}^{r}) \right\}^2, \quad (3.2)
\end{aligned}
$$

where e_{i1}^2 and e_{i2}^2 are prespecified weights. Intuitively, $M_N(\beta)$ and $M_n(\beta)$ measure the differences between two estimators of marginal density: the semiparametric estimator based on the attrition model and the nonparametric kernel estimator using either the first wave or the refreshment sample. Only with the true attrition parameters do the semiparametric estimators provide consistent estimates of the marginals with the objective function $M_{N,n}$ being close to zero. Our estimator $\widehat{\beta}$ is the minimizer such that $M_{N,n}$ is as close to zero as possible.

In (3.2), the weights e_{i1}^2 and e_{i2}^2 enable us to adaptively compare the differences between the two types of marginal density estimators. For example, it is well known that the performance of kernel density estimators is less satisfactory at the boundary due to the edge effect. Our simulation studies suggest that weighting, specifically trimming out data near the boundary, can potentially improve the estimation performance for two-wave data with a distribution that has a heavy tail. However, the advantage of weighting diminishes as the sample size increases. In addition, for distributions with light tails, such as the normal distribution, no weighting, with $e_{i1} = e_{i2} = 1$, gives better estimation performance. In practice, no weighting is recommended, in general, unless there is prior information on the distribution of the data or there is a preference for which regions to focus on when comparing these marginal density estimators.

3.3. Asymptotic theory

To establish our asymptotic results, we need the following conditions.

(A1) Let $S = \{(y_1, y_2) : f(y_1, y_2) > 0\}$ be the compact support of (Y_1, Y_2). Assume $S = [-t, t] \times [-u, u]$, and the support of $f(y_1, y_2 \mid W = 1)$ coincides with S.

(A2) The densities $f(y_1, y_2)$ and $f(y_1, y_2 \mid W = 1)$ are uniformly continuous and bounded away from zero on S.

(A3) The parameters $\beta = (\beta_0, \beta_1, \beta_2)$ belong to a compact set Θ, and without loss of generality, $\beta_0 \in [-b_0, b_0]$, $\beta_1 \in [-b_1, b_1]$, and $\beta_2 \in [-b_2, b_2]$.

(A4) The kernel function $K(y)$ is a probability density function and satisfies $|y|^{2+\delta} K(y) \to 0$ as $|y| \to +\infty$, for some $\delta > 0$.

(A5) For the two-dimensional kernel, the bandwidth $H = hI_2$, where I_2 is a 2×2 identity matrix and $h \to 0$ and $n_c h^4 / \log(n_c) \to +\infty$ as $n_c \to +\infty$.

(A6) The bandwidths h_1 and h_2 satisfy $h_1 \to 0$ and $h_2 \to 0$, and $(Nh_1^2)^{-1} \log N \to 0$ as $N \to +\infty$ and $(nh_2^2)^{-1} \log n \to 0$ as $n \to +\infty$, where N is the panel size, and n is the refreshment sample size.

Conditions (A1)–(A6) are common in the literature. Conditions similar to (A1)–(A3) are also considered in Hirano et al. (2001) and Bhattacharya (2008). Conditions (A4)–(A6) are needed to ensure the uniform consistency of the univariate and bivariate kernel density estimators, as in Devroye and Wagner (1980).

Let $\beta^0 = (\beta_0, \beta_1, \beta_2)$ be the true attrition parameters. Theorem 1 shows that β^0 is identified based on the marginal distributions of Y_1 and Y_2. Our main theoretical results are presented in Theorems 2 and 3, which establish the consistency and asymptotic normality, respectively, of the proposed semiparametric estimator.

Lemma 1. *Suppose conditions* (A1) *and* (A2) *are satisfied. Then for almost all* $(y_1, y_2) \in S$*, there is a unique set of parameters* $(\beta_0, \beta_1, \beta_2)$ *satisfying*

$$\int \frac{P(W=1)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1) dy_2 = f_1(y_1),$$
$$\int \frac{P(W=1)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1) dy_1 = f_2(y_2). \tag{3.3}$$

Proof of Lemma 1. The proof follows from Theorem 1 of Hirano et al. (2001).

Theorem 1 (Identifiability). *Under assumptions* (A1)–(A2)*, the two constraints in Equation* (3.3) *are uniquely satisfied by the true parameters* $\beta^0 = (\beta_0^0, \beta_1^0, \beta_2^0)$.

Proof of Theorem 1. Given the attrition model as $P(W = 1 \mid y_1, y_2) = \text{logistic}(\beta_0^0 + \beta_1^0 y_1 + \beta_2^0 y_2)$, it is sufficient to show that β^0 satisfies Equation (3.3), with $P(W = 1) f(y_1, y_2 \mid W = 1)/\text{logistic}(\beta_0^0 + \beta_1^0 y_1 + \beta_2^0 y_2) = f(y_1, y_2)$.

Theorem 2 (Consistency). *Under assumptions* (A1)–(A6)*, as* $n, N \to +\infty$*, the minimizer* $\widehat{\beta}$ *of* $M_{N,n}(\beta)$ *converges in probability to* β^0*, which is the unique minimizer of* $E\left\{f_1(Y_1 \mid \beta) - f_1(Y_1)\right\}^2 + E\left\{f_2(Y_2 \mid \beta) - f_2(Y_2)\right\}^2$.

The proof of Theorem 2 is presented in the Supplemental Material. It contains two main steps. First, $M_{N,n}(\beta)$ is shown to converge to its probability limit uniformly. Second, we show that this probability limit has a unique minimizer β^0. Then, the consistency follows from Theorem 5.7 of van der Vaart (2000).

Theorem 3 (Asymptotic Normality). *Suppose* $N/n \to r$*, for a constant* $r > 0$*. Under assumptions* (A1)–(A6)*, we have* $\sqrt{N}(\widehat{\beta} - \beta^0) \sim N(\mathbf{0}, \mathbf{V}^{-1}\mathbf{\Sigma}(\mathbf{V}^{-1})^T)$*, where* $\mathbf{V} = E\{\partial^2 M_N(\beta^0)/\partial\beta\partial\beta^T\} + E\{\partial^2 M_n(\beta^0)/\partial\beta\partial\beta^T\}$ *and* $\mathbf{\Sigma} = 4\mathbf{\Sigma}_1 + \mathbf{\Sigma}_{21} + 4r\mathbf{\Sigma}_{22} + 4\mathbf{\Sigma}_{cov}$*, defined in* (A6) *and* (A7)*, respectively, Supplementary Material.*

The asymptotic property of $\widehat{\beta}$ is evaluated using a Z-estimator by taking the derivative of $M_{N,n}(\beta)$. There are two parts in $M_{N,n}(\beta)$ from (3.2), namely, $M_N(\beta)$ and $M_n(\beta)$. In the proof included in the Supplemental Material, we tackle each part separately. Theorem 3 combines the asymptotic expansions of these two parts.

3.4. Hypothesis testing

The asymptotic theory developed in subsection 3.3 can be used to perform hypothesis testing for missing mechanisms by testing the attrition parameters β_1 and β_2 in the additive non-ignorable model. For MCAR, MAR, and MNAR, consider $H_0: \ \beta_1 = 0$ and $\beta_2 = 0$, $H_0: \ \beta_2 = 0$, and $H_0: \ \beta_2 \neq 0$, respectively. A Wald-type test statistic can be constructed based on the asymptotic normality of the semiparametric estimators $\widehat{\beta}_i$,

$$Z = \frac{\widehat{\beta}_i - \beta_{i0}}{SE_{\widehat{\beta}_i}} = \frac{\widehat{\beta}_i}{SE_{\widehat{\beta}_i}}, \quad \text{for } i = 1, 2, \tag{3.4}$$

where $SE_{\widehat{\beta}_i}$ are corresponding standard errors. The $100(1-\alpha)\%$ confidence interval can be defined as $\widehat{\beta}_i \pm z_{1-\alpha/2} SE_{\widehat{\beta}_i}$, for $i = 1, 2$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. The asymptotic theory in Theorem 3 gives the asymptotic formula for computing the standard errors. However, it requires both the true population density functions and the true attrition parameters, which are often unavailable in practice. Therefore, we propose using a bootstrap to approximate the standard errors numerically. The accuracy of this bootstrap SE is assessed numerically by comparing it to the empirical SE in the simulation studies. In addition, we compare the power functions of the test statistics defined in (3.4) with different standard errors.

4. Simulation Studies

This section evaluates the numerical performance of the proposed full likelihood and kernel-based semiparametric methods. Each simulation in this section includes 1,000 replications.

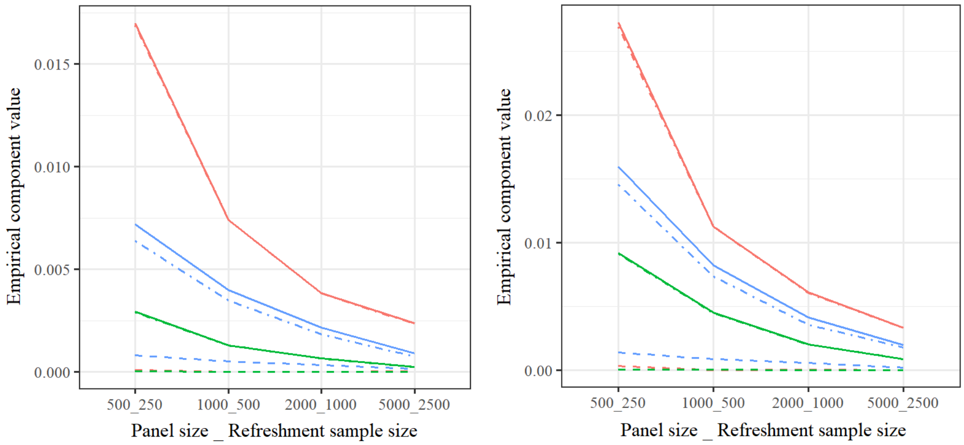
4.1. Comparison of three estimation methods

We first compare the finite-sample performance of the proposed full likelihood (or parametric) and semiparametric methods with that of Bhattacharya's conditional moment restriction (CMR) method. Data sets are generated from the bivariate normal and gamma-t distributions. The gamma-t distribution is used to understand the effect of a model misspecification.

Two-wave data (Y_1, Y_2) are generated independently from a bivariate normal distribution with mean $\mathbf{0}$, marginal variances 10, and correlation coefficient 0.5. The true attrition follows a logistic regression with attrition parameters of $\beta_0 = 0$, $\beta_1 = 0.3$, and $\beta_2 = 0.4$. Three methods are applied to obtain estimates of the attrition parameters. Figure 1 compares the finite-sample performance in terms of the empirical squared bias, variance, and MSE for $\widehat{\beta}_1$ and $\widehat{\beta}_2$. The x-axis shows panel size and refreshment sample size combinations, with both sample sizes increasing along the x-axis. Figure 1 clearly shows that the MSEs of both the parametric and the semiparametric methods decrease as the sample sizes increase, corroborating the asymptotic results. In addition, the parametric and semiparametric methods outperform the CMR method, where the latter has the largest MSE in all sample size combinations. In particular, for a panel size of 5,000 and a refreshment size of 2500, the parametric estimator of β_1 has about one-third the variance of the semiparametric estimator, which, in turn, has nearly one-third the variance of the CMR estimator. Due to the attrition in the second wave, the variances of $\widehat{\beta}_2$ are larger for all three methods. The parametric estimator of β_2 has about half the variance of the semiparametric estimator, which, in turn, has about half the variance of the CMR estimator.

To generate non-normal data, we consider the marginal distributions of the first and second waves as $Gamma(3, 2)$ and $t(6)$, respectively. To make the distributions comparable with the previous bivariate normal case, we shift the $Gamma$ distribution to a center at zero, and the t distribution is scaled by three. Copulas are used to create a non-normal joint density with the given marginals and a correlation coefficient of 0.5 (Yan, 2007). As a result, the joint distribution centers at zero, and the $Gamma$ marginal has a variance of 12, and the t-distribution has a variance of 13.5. Compared with the bivariate normal distribution, this distribution has the same zero means and slightly larger marginal variances.

For the performance of $\widehat{\beta}_1$, Figure 2 shows that the parametric method performs better in terms of the MSE. However, as the sample size increases, the



(a) Finite-sample performance of $\widehat{\beta}_1$ (b) Finite-sample performance of $\widehat{\beta}_2$

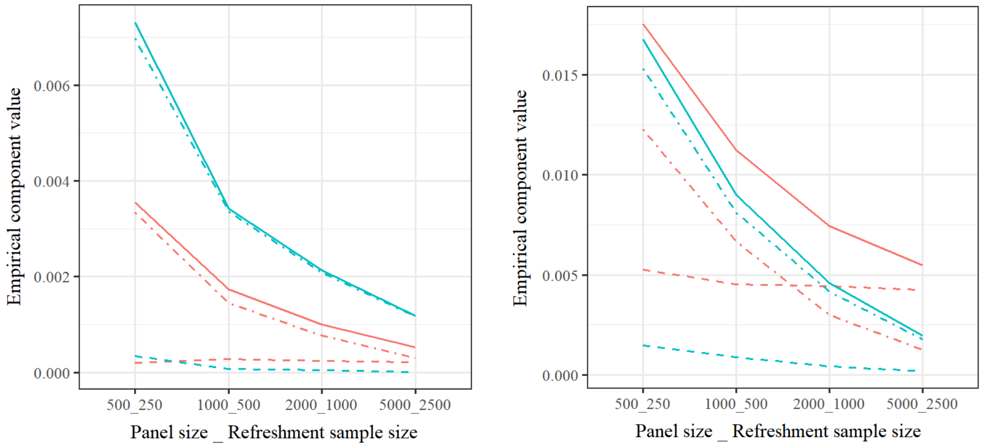
Figure 1. Comparison of finite-sample performance of three estimation methods: parametric (triangle), semiparametric (circle), and CMR (x) for bivariate normal responses. The dash, dot-dash, and solid lines represent the empirical squared bias, variance, and MSE, respectively.

Table 2. Gamma-t population. Empirical squared bias, variance, and MSE of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for parametric and semiparametric methods with a panel size of 5,000 and refreshment sample size of 2,500.

	Squared Bias (10^{-3})		Variance (10^{-3})		MSE (10^{-3})	
	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
Semiparametric	0.0087	0.5082	1.1457	1.5718	1.1544	2.0801
Parametric	0.2161	4.2440	0.3026	1.2451	0.5183	5.4891

parametric method has a nondecreasing bias, whereas the semiparametric method has a decreasing bias. The variance of the semiparametric estimator $\widehat{\beta}_1$ is still larger than that of the parametric estimator. However, for $\widehat{\beta}_2$, the parametric method gives a noticeably larger bias, and leads to a larger MSE than does the semiparametric method. The same observations are evident in Table 2, which reports the empirical squared bias, variance, and MSE of the parametric and semiparametric estimators for a panel size of 5,000 and a refreshment sample size of 2,500.

In the bivariate normal setting, our proposed parametric and semiparametric methods outperform the CMR method. When the joint distribution is specified correctly, the parametric method outperforms the other two methods. However, when the distribution is misspecified, there is bias in the parametric estimator, whereas the semiparametric estimator, which is free of distributional assumptions, yields a consistent performance in the presence of non-normal populations.



(a) Finite-sample performance of $\widehat{\beta}_1$ (b) Finite-sample performance of $\widehat{\beta}_2$

Figure 2. Comparison of finite-sample performance with gamma-t responses: parametric (triangle) and semiparametric (circle). For both methods, the dash, dot-dash, and solid lines represent the empirical squared bias, variance, and MSE, respectively.

4.2. Effect of weighting

As discussed in section 3.2, weight assignments allow us to prioritize the comparison of the density function estimates over different regions of interest. To investigate the effect of weights, we generate data from three distinct distributions: a bivariate normal, as in subsection 4.1, a uniform distribution, and a beta distribution. For the uniform and beta distributions, the two-wave data Y_1 and Y_2 are independent, and both follow either a $Unif(-\sqrt{30}, \sqrt{30})$ or a scaled beta distribution with location and scale parameters 0.5 and 0.5, respectively, and a minimum of $-2\sqrt{5}$ and a maximum of $2\sqrt{5}$. In all three distributions, the two-wave data have the same marginal mean of zero and variance of 10. We consider two weighting strategies, $e_{1,i1} = e_{1,i2} = 1$, $e_{2,i1} = I(q_{1,0.05} \leq Y_{i1} \leq q_{1,0.95})$ and $e_{2,i2} = I(q_{2,0.05} \leq Y_{i2} \leq q_{2,0.95})$. Here, $q_{1,\alpha}$ and $q_{2,\alpha}$ are the αth sample quantiles for Y_1 and Y_2, respectively. The first set $e_{1,i1}$, $e_{1,i2}$ imposes no weighting, and the second set $e_{2,i1}$, $e_{2,i2}$ considers only the middle 90% of the data.

Table 3 reports the empirical squared bias, variance, and MSE of $\hat{\beta}_1$ and $\hat{\beta}_2$ for the proposed semiparametric estimator under the two weighting schemes. For the normal distribution, the estimators without a weighting (e_1) perform better, with smaller MSEs for both sample size combinations. However, for both the uniform and the beta distributions, the estimators with the weighting (e_2) perform slightly better. This indicates that weighting can be useful in mitigating the edge effect of a kernel density estimation, especially for distributions with heavy tails. However, the advantage of weighting diminishes as the sample size increases.

Table 3. Empirical squared bias, variance, and MSE of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for semiparametric methods with two weights, e_1: no weight, and e_2: a weight that uses only 90% of the data under different distributions.

				Squared Bias (10^{-3})		Variance (10^{-3})		MSE (10^{-3})	
Distribution	n_1	n_2		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
Normal	100	50	e_1	0.82	1.26	33.43	70.53	34.25	71.79
			e_2	1.10	1.63	42.93	84.19	44.03	85.83
	500	250	e_1	0.94	1.59	5.01	10.89	5.95	12.48
			e_2	0.80	1.40	6.39	14.57	7.19	15.97
Uniform	100	50	e_1	2.80	5.26	8.88	36.43	11.69	41.69
			e_2	1.85	3.02	9.95	33.63	11.80	36.66
	500	250	e_1	2.34	4.51	3.91	5.02	6.25	9.53
			e_2	2.19	3.12	4.02	5.27	6.22	8.39
Beta	100	50	e_1	6.07	9.12	9.46	38.60	15.53	47.72
			e_2	5.65	8.41	9.47	39.05	15.12	47.47
	500	250	e_1	5.59	7.34	1.99	3.52	7.59	10.85
			e_2	4.95	6.42	2.13	3.78	7.08	10.19

4.3. Bootstrapping in applications

We evaluate the numerical performance of the proposed Wald test using three approaches to calculate the standard error: empirical SE (ESE), asymptotic SE (ASE), and bootstrap SE (BSE). The ESEs are calculated from 1,000 simulation replications, and serve as a benchmark for comparison, but are not available in practice. The ASEs are based on the asymptotic variance in Theorem 3, which requires knowledge of the true parameter values and population density functions, making it often impractical. Thus, we propose using a bootstrap as an alternative to approximate the standard errors. We compare the performance of these approaches based on the power of the corresponding test statistics.

In the bootstrap method, 500 bootstrap samples are created. Each bootstrap sample consists of a bootstrapped panel and a bootstrapped refreshment sample, which are bootstrap samples from the original panel and the refreshment sample, respectively. The semiparametric method is applied to each bootstrap sample to estimate the attrition parameters, and the standard deviation of these 500 estimates is the BSE.

A total of 200 samples with panel size 5,000 and refreshment size 2,500 are drawn from a bivariate normal population, each with the marginal mean zero, variance 10, and correlation coefficient 0.5. For each sample, we perform the Wald test at the significance level of $\alpha = 0.05$. In the Wald test statistic, three different SEs are considered. The proportion of rejecting the null hypotheses in the 200 replications is calculated as the empirical power for each method, and is evaluated at $(0, 0.05, 0.1, 0.2, 0.3)$ for β_1 and $(0, 0.05, 0.1, 0.13, 0.2, 0.4)$ for β_2.

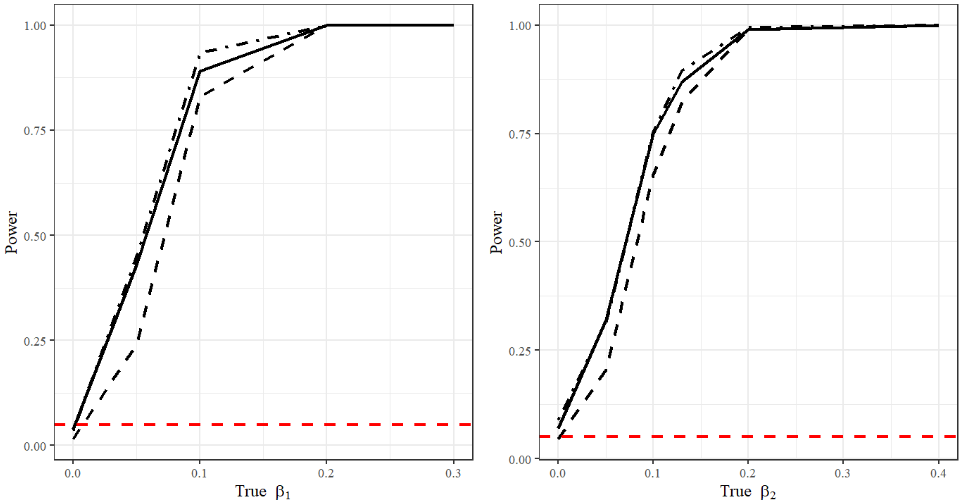


Figure 3. Power function comparison. The solid, dash, and dot-dash lines represent the power functions based on the bootstrap SE (BSE), asymptotic SE (ASE), and empirical SE (ESE), respectively. The dash line at the bottom indicates the significance level, 0.05.

Figure 3 gives the power functions based on the BSE (solid), ASE (dash), and ESE (dot-dash). For all three methods, the power is close to the significance level of 0.05 when $\beta_i = 0$. In addition, the power increases quickly to one as the true value of β_i moves away from the hypothesized value of zero for all three methods, indicating that the proposed Wald test works reasonably well. More importantly, the power functions based on the BSE and ESE are close to each other, and both have overall higher power than those based on the ASE. This shows that the Wald test based on the bootstrap SE works reasonably well.

In addition, the 95% confidence intervals for β_1 and β_2 are constructed based on the ASE and BSE. Table 4 reports the empirical coverage probabilities of these confidence intervals for different choices of β_1 and β_2. Overall, the confidence intervals based on the BSE have empirical coverage probabilities closer to the nominal level of 95%. In contrast, the confidence intervals based on the ASE are more conservative, with empirical coverage probabilities higher than 95%.

5. Netherlands Mobility Panel

The Netherlands Institute for Transport Policy Analysis has conducted the Netherlands Mobility Panel (NMP) since 2013, a multiple-wave longitudinal study aimed at understanding changes in travel behavior over time. Detailed information can be found in Hoogendoorn-Lanser, Schaap and OldeKalter (2015).

The NMP samples households as survey units and collects travel information by distributing questionnaires to members in each household. The NMP conducted the initial and second wave of the data survey in 2013 and

Table 4. Coverage probabilities of 95% confidence intervals for β_1 and β_2 based on 200 replications. The standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are computed using the asymptotic formula and bootstrapping. The panel size is 5,000, and the refreshment sample size is 2,500. The coverage probabilities are calculated for different true values of attrition parameters β_1 and β_2.

		Asymptotic Formula		Boostrap	
β_1	β_2	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
0	0	1	0.995	0.985	0.945
0.05	0.05	1	1	0.975	0.955
0.1	0.1	1	1	0.990	0.965
0.2	0.2	1	1	0.985	0.940
0.3	0.4	1	0.980	0.990	0.960

2014, respectively. The database consists of three components: household data, personal data, and individual travel diary data. Based on the household data, there were 3,572 households in the initial wave, and 4,685 households in the second wave. In the first wave, 2,380 households provided household information and travel diary data. Of the 2,380 complete cases, 1,685 households continued to report their travel behaviors during the second wave of the data collection, while the remaining 695 households did not respond. A refreshment sample of 1,382 households with household information and travel diary data was identified and collected simultaneously as the second wave. These sets of 2380, 695, and 1382 households represent the complete set, incomplete set, and refreshment sample, respectively.

The NMP examines travel behavior over time, and various studies have analyzed these data to gain insights into such behaviors. For instance, Kroesen and Goulias (2016) investigates the relationship between attitudes and travel behaviors using the complete data set. Hoogendoorn-Lanser, Schaap and Old-eKalter (2015) estimates the nonresponse bias by modeling the nonresponse behavior using a logistic regression and a MAR assumption. Puello and Geurs (2016) explores the effects of nonrandom attrition on mobility rates using trip diary data. Their analysis assumed MAR assumption and attrition was evaluated only through observed demographic data.

The aforementioned studies assume MCAR or MAR in their analyses of the NMP data. In contrast, we relax the missing mechanism assumption and consider MNAR. We use the refreshment sample to estimate and draw inferences about the MNAR attrition parameters, which yields insights into the true missing mechanism. Specifically, we focus on total travel time as the variable of interest and investigate whether the missing mechanism is related to this variable. The travel diary records all trips made by each household over three days, and we calculate the total travel time by summing the travel times and rescaling the sum using a natural log transformation.

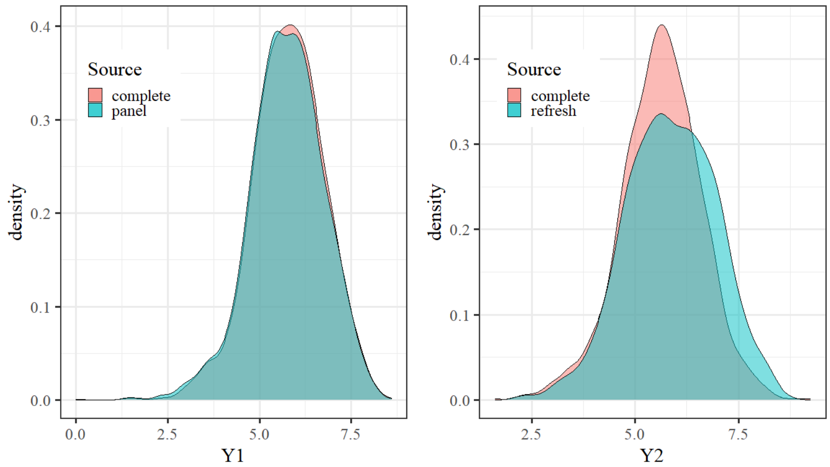


Figure 4. Marginal density comparison of NMP on the first and second wave.

Figure 4 compares the marginal densities of the log-transformed total travel time on each wave. Here, Y_1 and Y_2 are the total travel times on the natural log scale at the initial and second waves, respectively. In the left panel of Figure 4, the estimated marginal density of Y_1 based on the complete set is shown in gray, while the one based on the full panel Y_1 is in dark gray. In the right panel, the estimated marginal density of Y_2 based on the complete set is shown in gray, and the density based on the refreshment sample is shown in dark gray. The estimated marginal densities of Y_1 and Y_2 based on the complete set can be biased due to missingness in the data. In contrast, the full panel Y_1 and the refreshment sample provide more accurate estimates for the true marginal densities.

We consider three possible attrition models corresponding to the three missing mechanisms, MCAR, MAR, and MNAR. Let W_i denote the missingness (attrition) indicator for the ith subject with $W_i = 1$ if Y_2 is observed for subject i and $W_i = 0$ otherwise. We assume an additive logistic model for the probability of $W_i = 1$ as $\pi_{\text{MNAR}} = P(W_i = 1 \mid y_1, y_2, \beta) = \text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)$. This reduces to MCAR when $\beta_1 = \beta_2 = 0$ and to MAR when $\beta_2 = 0$.

Table 5 gives the estimation results of the attrition parameters and their 95% confidence intervals under the three missing mechanisms. Under MNAR, neither of the confidence intervals for β_1 and β_2 contain zero, indicating strong evidence that the missingness is related to Y_1 and Y_2. Therefore, neither MCAR nor MAR are adequate assumptions for the NMP. In addition, the positive estimate of β_1 indicates that the probability of being observed in the second wave increases as the value of the total travel time in the first wave increases, and the negative estimate of β_2 indicates that the probability of being observed decreases with the value of total travel time in the second wave. This is also consistent with the observation in Figure 4 that the complete set has a density leaning toward lower values of Y_2,

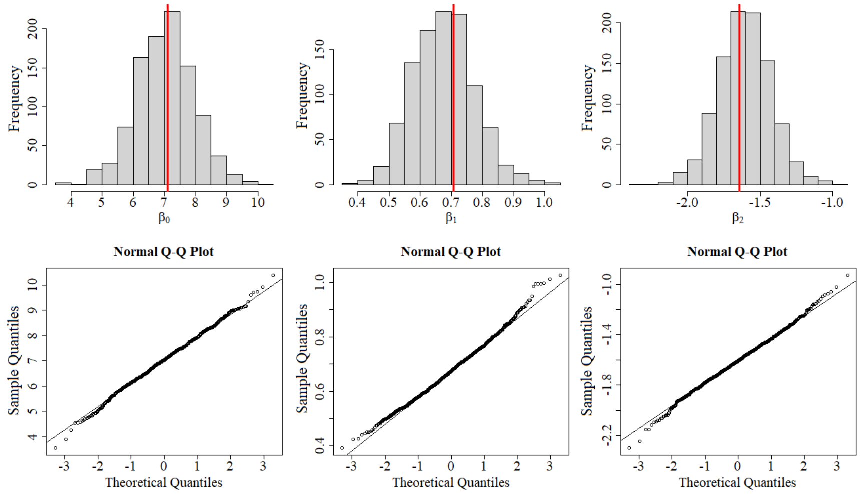


Figure 5. Sampling distributions of bootstrapped semiparametric estimators in NMP application.

Table 5. Point estimates and 95% confidence intervals for attrition parameters for NMP data.

Attrition model	MCAR	MAR	MNAR
$logit(\pi) =$	β_0	$\beta_0 + \beta_1 y_1$	$\beta_0 + \beta_1 y_1 + \beta_2 y_2$
$\widehat{\beta}_0$	0.89 (0.80, 0.97)	0.03 (-0.47, 0.54)	7.11 (5.09, 8.91)
$\widehat{\beta}_1$		0.15 (0.06, 0.24)	0.71 (0.50, 0.88)
$\widehat{\beta}_2$			-1.64 (-1.97, -1.25)

compared with the marginal density from the refreshment sample. Under MNAR, the 95% confidence intervals are constructed using bootstrapping. Figure 5 plots the sampling distributions of the bootstrapped semiparametric estimators. The vertical lines represent the point estimates from the original data.

6. Conclusion

We extend the method of Hirano et al. (2001) for identifying and estimating the non-ignorable attrition mechanism for binary responses to continuous responses, using a refreshment sample in two-wave panel data. The introduction of refreshment samples into missing data analysis enables researchers to test the missing mechanism assumption. The proposed full likelihood method relies on the correct specification of the underlying population and attrition mechanism, which is impractical in practice. The kernel-based semiparametric method is the primary approach we propose to reduce the unavoidable bias due to model misspecification. We show the consistency and asymptotic normality of the additive attrition estimators in the semiparametric model.

Current methods are limited to data with only two waves. Extending our methods to multi-wave data is challenging, owing to the curse of dimensionality in a multivariate nonparametric density estimation. However, our method can be extended to a more flexible attrition model by using a nonparametric or additive link function. These generalizations increase the robustness of our method and enable its application to data with a more general structure, which is worth future investigation. Furthermore, the current model setup does not consider any covariates. We extend the proposed method in the Supplemental Material to include binary or discrete covariates. However, investigating the case with more general covariates is left to future research.

Supplementary Material

The online Supplementary Material includes an extension to binary covariates, an additional simulation, necessary lemmas, and detailed proofs.

Acknowledgments

The research reported here was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Number R01DK132385. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. In addition, Wang's research was partially supported by a UIC LAS Mid-Career Award. Qu's research was supported by the National Science Foundation (DMS-1952406)

References

Asigbee, F. M., Whitney, S. D. and Peterson, C. E. (2018). The link between nutrition and physical activity in increasing academic achievement. *Journal of School Health* **88**, 407–415.

Bhattacharya, D. (2008). Inference in panel data models under attrition caused by unobservables. *Journal of Econometrics* **144**, 430–446.

Deng, Y. (2012). *Modeling Missing Data in Panel Studies with Multiple Refreshment Samples*. Ph.D. thesis. Duke University.

Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y. and Zheng, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. *Statistical Science* **28**, 238–256.

Devroye, L. P. and Wagner, T. J. (1980). The strong uniform consistency of kernel density estimates. In *Multivariate Analysis V: Proceedings of the 5th International Symposium on Multivariate Analysis*, 59–77.

d'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics* **154**, 1–15.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008). *Longitudinal Data Analysis*. CRC Press.

Hirano, K., Imbens, G. W., Ridder, G. and Rubin, D. B. (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica* **69**, 1645–1659.

Hoogendoorn-Lanser, S., Schaap, N. T. and OldeKalter, M.-J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation Research Procedia* **11**, 311–329.

Ingels, S. J., Pratt, D. J., Alexander, C. P., Jewell, D. M., Lauff, E., Mattox, T. L. et al. (2014). *Education Longitudinal Study of 2002 (ELS: 2002) Third Follow-up Data File Documentation.* National Center for Education Statistics.

Kim, S. (2009). *Sample Attrition in the Presence of Population Attrition.* Technical report. University of Washington.

Kroesen, M. and Goulias, K. G. (2016). Modeling activity-travel behavior dynamics with panel data: The state-of-the-art. *European Journal of Transport and Infrastructure Research* **16**, 633–637.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data.* John Wiley & Sons.

Miao, W., Ding, P. and Geng, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association* **111**, 1673–1683.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* Springer, New York.

Morikawa, K. and Kim, J. K. (2021). Semiparametric optimal estimation with nonignorable nonresponse data. *The Annals of Statistics* **49**, 2991–3014.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business & Economic Statistics* **21**, 43–52.

Puello, L. L. P. and Geurs, K. T. (2016). Integration of unobserved effects in generalised transport access costs of cycling to railway stations. *European Journal of Transport and Infrastructure Research* **16**, 385–405.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* **128**, 301–323.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons.

Si, Y., Reiter, J. P. and Hillygus, D. S. (2015). Semi-parametric selection models for potentially non-ignorable attrition in panel studies with refreshment samples. *Political Analysis* **23**, 92–112.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Chapman and Hall/CRC.

Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **172**, 659–687.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

Wang, S., Shao, J. and Kim, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* **24**, 1097–1116.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software* **21**, 1–21.

(Received June 2022; accepted April 2023)