Statistica Sinica Preprint No: SS-2022-0214	
Title	Semiparametric Estimation of Non-ignorable Missingness With Refreshment Sample
Manuscript ID	SS-2022-0214
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0214
Complete List of Authors	Jianfei Zheng, Jing Wang, Lan Xue and Annie Qu
Corresponding Authors	Lan Xue
E-mails	xuel@science.oregonstate.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

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: Missing data commonly arises in longitudinal data analysis and imposes methodological challenges in providing unbiased estimation and statistical inference due to informative missingness. 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 MCAR and MAR. 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. If the specification of the joint distribution is unavailable, a semiparametric method is proposed to estimate the attrition parameters with marginal density estimates obtained using an additional refreshment sample. We derive asymptotic properties of the semiparametric estimators and illustrate the numerical performances with simulations. Inference based on bootstrapping is proposed and assessed through simulations. 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 have been widely used in many 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 et al., 2015) has been conducting the Netherlands Mobility Panel since 2013. This panel currently involves two waves of data collection, with an initial wave consisting of 2380 households. For the second wave, only 1685 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 responses. Therefore, understanding the missing mechanism is crucial in 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). In MCAR, the missingness is assumed to be independent of all the variables in the data including, both observed and missing, while MAR allows the missing mechanism to depend on variables that are

always observed. On the other hand, MNAR further relaxes the assumption for the missing mechanism and assumes the missingness to depend on both observed and unobserved variables. Different statistical methods have been developed to allow valid estimations and inferences under these missingness assumptions.

Unfortunately, partially observed panel data alone cannot distinguish among various missing mechanisms, and the assumptions mentioned above for missingness 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 as the panel data alone is often insufficient to make inferences about populations (Rubin, 1976, 2004; Hirano et al., 2001; Fitzmaurice et al., 2008). Miao et al. (2016) provided sufficient conditions for model identifiability when the response follows a normal or a normal mixture distribution. d'Haultfoeuille (2010), under a completeness assumption, and Wang et al. (2014), based on the generalized method of moments, established 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) provided a sufficient condition for its identifiability without instrumental variable assumption.

Hirano et al. (2001) first explored the use of 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 et al., 2018) and the National Educational Longitudinal Study (Ingels et al., 2014), refill their samples once or multiple times during a study. The Netherlands Mobility Panel completed its initial survey data in 2013, a follow-up survey was administered in 2014, and a refreshment sample was considered and incorporated.

Refreshment samples provide an inexpensive way to improve the quality of longitudinal data. Methods have been developed to estimate the attrition process with a refreshment sample. Hirano et al. (2001) proposed 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 provided the fundamental identification theory and developed an estimation procedure for a two-wave binary response. Nevo (2003) explored the refreshment sample to compute sampling weights for adjustment of missingness. The weights are constructed so that the moments of the weighted data match those observed in the refreshment sample. Bhattacharya (2008) converted Hirano's fundamental identification theory into conditional moment restrictions. A set of nonparametric regressions

with B-splines were used to construct the objective function for 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, where a fully Bayesian approach is implemented with Markov chain Monte Carlo estimation. Similarly, Si et al. (2015) presented a semiparametric additive non-ignorable model to analyze multivariate categorical responses in a two-wave panel with one refreshment sample, which adopted the additive non-ignorable model for the attrition process and modeled the multinomial survey responses with a Dirichlet process mixture.

This paper proposes two new approaches that handle 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 are made through maximum likelihood estimators. Adaptive Gaussian quadrature is used to overcome the integration difficulty introduced by the missing data in the construction of the likelihood. A second method is a semiparametric approach utilizing the kernel density estimator as the nonparametric component, and the additive non-ignorable attrition model (Hirano et al., 2001) is adopted 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 correctly specified, the full-likelihood approach provides the most efficient estimators and acts as a benchmark for comparing different methods of analyzing MNAR data 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 distributional specification and provides consistent inferences for the attrition process under different population conditions. Simulation results also support the finding that the kernel density-based semiparametric estimators have better numerical performance than the method proposed by Bhattacharya (2008).

The first contribution of this paper 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 the 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) and Deng et al. (2013), the semiparametric estimator is shown to be consistent and asymptotically normal, and inference tools are developed to test the MCAR and MAR assumptions based on asymptotic formulas and bootstrapping methods. The proposed methods differ fundamen-

tally from those designed for binary data (Hirano et al., 2001; Deng et al., 2013), as the distribution of binary data can be characterized with a few parameters, and the estimation procedure only involves 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 estimation and inference of the attribution parameters. Extensive simulation results are given in Section 4. An application using the Netherlands Mobility Panel is illustrated in Section 5. Finally, Section 6 summarizes the present research and discusses future research possibilities.

2. Refreshment Sample and Models

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

The refreshment sample is an additional independent random sample from the population during follow-up waves when attrition starts to occur. Suppose $\{\mathbf{Y}_i = (Y_{i1}, Y_{i2})\}_{i=1}^N$ are 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, while 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, 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, which is 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 the MAR models as special cases. In particular, it leads to the MCAR model if both κ_1 and κ_2 are 0, and to the MAR model if only κ_2 is 0. When κ_2 is nonzero, the data are MNAR. Therefore, it provides an approach to test for MCAR or MAR mechanisms through testing for nonzero κ's. This model still has an untestable as-

sumption that the missingness depends additively on the responses, without any interactions. According to Hirano et al. (2001), this is the weakest assumption that can be identifiable and estimable using a refreshment sample.

When both Y_{i1} and Y_{i2} are binary, Hirano et al. (2001) provided two fundamental identification constraints for the attrition parameters and proposed to estimate the attrition parameters using the method of moments. An implementation of the additive non-ignorable model was not given for continuous responses. We aim to extend Hirano et al. (2001)'s approach 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 to describe the missing mechanism (Rubin, 1976; Hirano et al., 2001; Nevo, 2003; Bhattacharya, 2008; Kim, 2009; Little and Rubin, 2019). In addition to the logistic regression, the probit model is another popular choice for the missing mechanism. Miao et al. (2016) provided sufficient conditions for the probit model to be identified when the response variable follows a normal

or normal mixture distribution without a refreshment sample. Our proposed method can be directly extended to other parametric attrition models, including the probit model. In addition, our proposed method can 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 aim to develop methods to handle two-wave MNAR data with continuous responses instead of binary ones. In this section, we introduce two new methods which 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 attrition parameters based on Hirano's constraints. The asymptotic theory of the semiparametric estimator is developed in subsection 3.3. We also 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 nor-

mal. 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. To be specific, 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, as it requires the integration of a joint density for each incomplete data point, and there is no closed-form solution. To address this, we propose to use adaptive Gaussian-Hermite quadrature (Skrondal and Rabe-Hesketh, 2004; Rabe-Hesketh et al., 2005; Skrondal and Rabe-Hesketh, 2009) for numerical approximation. The Gaussian-Hermite quadrature is a commonly used technique for generalized lin-

ear mixed models (Molenberghs and Verbeke, 2005).

In the parametric approach, the refreshment sample helps to identify parameters θ and β in the observed likelihood $L(\theta, \beta)$. Miao et al. (2016) provided sufficient identifiable conditions for normal response or normal mixture in a probit model. Without refreshment samples, the model parameters are generally 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 attrition model are correctly specified. However, misspecification of either the population distribution or the attrition model can lead to biased estimation and inference. In the next section, we introduce a semiparametric method that does not require the 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 correctly specified, one can reconstruct the unobserved joint density

$f(y_1, y_2)$ from the observed counterpart $f(y_1, y_2 \mid W = 1)$ by $f(y_1, y_2) = \frac{P(W=1)}{P(W=1|y_1,y_2)} f(y_1, y_2 \mid W = 1)$. As a result, for marginal densities, one has

$$\begin{aligned} \int \frac{P(W=1)}{P(W=1|y_1,y_2)} f(y_1, y_2 \mid W=1) dy_2 &= f_1(y_1), \\ \int \frac{P(W=1)}{P(W=1|y_1,y_2)} f(y_1, y_2 \mid W=1) dy_1 &= f_2(y_2), \end{aligned} \tag{3.3}$$

where f_1 and f_2 are marginal densities for Y_1 and Y_2. 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 $\widehat{f}_H(\mathbf{y} \mid W = 1) = \frac{1}{n_c} \sum_{i=1}^{n_c} K_H(\mathbf{y} - \mathbf{Y}_i)$, where $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$, $i = 1, 2, ..., n_c$ 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 consistently estimated 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 by $\widetilde{f}_1(y_1 \mid \beta) = \int \widetilde{f}(y_1, y_2 \mid \beta) dy_2$. For a given y_2, $\widetilde{f}_2(y_2 \mid \beta)$ can be defined similarly. Due to missingness, the refreshment sample, instead of data observed in the second wave, should be used 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 correctly specified.

The marginal densities on the right-hand side of Equation (3.3) 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 with $\widehat{f}_1(y_1) = \frac{1}{N}\sum_{i=1}^N K_{h_1}(y_1 - Y_{i1})$, and $\widehat{f}_2(y_2) = \frac{1}{n}\sum_{i=1}^n K_{h_2}(y_2 - Y_{i2}^r)$, 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, the plug-in method is used to select the bandwidths in the kernel density estimators, which is implemented with 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

$$\begin{aligned} & M_{N,n}(\beta) = M_N(\beta) + M_n(\beta) \\ = \; & \frac{1}{N}\sum_{i=1}^{N} e_{i1}^2 \left[\widetilde{f}_1(Y_{i1} \mid \beta) - \widehat{f}_1(Y_{i1})\right]^2 + \frac{1}{n}\sum_{i=1}^{n} e_{i2}^2 \left[\widetilde{f}_2(Y_{i2}^r \mid \beta) - \widehat{f}_2(Y_{i2}^r)\right]^2 \end{aligned} \tag{3.4}$$

where e_{i1}^2 and e_{i2}^2 are pre-specified 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, 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 (3.4), the weights e_{i1}^2 and e_{i2}^2 enable us to adaptively compare the differences between 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 weighing diminishes as the sample 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 a preference to focus on which regions to compare 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 0 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 2-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 commonly used 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 uniform consistency of univariate and bivariate kernel density estimators as in Devroye and Wagner (1980).

Let $\beta^0 = (\beta_0, \beta_1, \beta_2)$ be 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 consistency and asymptotic normality of the proposed semiparametric estimator, respectively.

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*

$$\begin{aligned} \int \frac{P(W=1)}{\textit{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1) dy_2 &= f_1(y_1), \\ \int \frac{P(W=1)}{\textit{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1) dy_1 &= f_2(y_2). \end{aligned} \tag{3.5}$$

Proof of Lemma 1. It follows from Theorem 1 of Hirano et al. (2001). □

Theorem 1. *(Identifiability) Under assumptions (A1)-(A2), the two constraints in Equation (3.5) 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 (3.5) with $\frac{P(W=1)}{\text{logistic}(\beta_0^0+\beta_1^0 y_1+\beta_2^0 y_2)} f(y_1, y_2 \mid W = 1) = f(y_1, y_2)$. $\square$

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), one has $\sqrt{N}\left(\widehat{\beta} - \beta^0\right) \sim N\left(\mathbf{0}, \mathbf{V}^{-1}\mathbf{\Sigma}\left(\mathbf{V}^{-1}\right)^T\right)$, where $\mathbf{V} = E\left[\frac{\partial^2}{\partial\beta\partial\beta^T} M_N\left(\beta^0\right)\right] + E\left[\frac{\partial^2}{\partial\beta\partial\beta^T} M_n\left(\beta^0\right)\right]$, and $\mathbf{\Sigma} = 4\mathbf{\Sigma}_1 + \mathbf{\Sigma}_{21} + 4r\mathbf{\Sigma}_{22} + 4\mathbf{\Sigma}_{cov}$ defined in (A6) and (A7) of supplemental material.*

The asymptotic property of $\widehat{\beta}$ is evaluated through a Z-estimator by taking the derivative of $M_{N,n}(\beta)$. There are two parts in $M_{N,n}(\beta)$ from (3.4), namely $M_N(\beta)$ and $M_n(\beta)$. In the proof included in the supplemental material, we tackle each part separately, and Theorem 3 follows by combining 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 of missing mechanisms, which can be accomplished by testing 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 $\widehat{\beta}_i$,

$$Z = \frac{\widehat{\beta}_i - \beta_{i0}}{SE_{\widehat{\beta}_i}} = \frac{\widehat{\beta}_i}{SE_{\widehat{\beta}_i}}, \quad \text{for } i = 1, 2, \tag{3.6}$$

where $SE_{\widehat{\beta}_i}$s 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, often unavailable in practice. Therefore we propose to use the bootstrap to approximate 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, the comparison is also made through power functions of the test statistics defined in (3.6) 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 1000 replications.

4.1 Comparison of Three Estimation Methods

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

Two-wave data (Y_1, Y_2) are independently generated from the bivariate normal with a mean of $\mathbf{0}$, marginal variances of 10, and correlation coefficient of 0.5. The true attrition follows a logistic regression with attrition parameters of $\beta_0 = 0$, $\beta_1 = 0.3$, $\beta_2 = 0.4$. Three methods are applied to obtain estimates of attrition parameters. Figure 1 compares the finite-sample performances in terms of empirical squared bias, variance, and MSE for $\widehat{\beta}_1$ and $\widehat{\beta}_2$, respectively. The x-axis gives different panel size and refreshment sample size combinations, with both sample sizes increasing along the x-axis. Figure 1 clearly shows that MSEs of both the parametric and semiparametric methods decrease as the sam-

ple sizes increase, which corroborates with the asymptotic results. In addition, the parametric and the semiparametric methods outperform the CMR method, with CMR having the largest MSE in all sample size combinations. In particular, for a panel size of 5000 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 one, 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 the second waves as $Gamma(3, 2)$ and $t(6)$, respectively. To make the distributions comparable to the previous bivariate normal case, we shift the $Gamma$ distribution to a center at 0, and the t distribution is scaled by 3. Copulas are used to create a non-normal joint density with 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, while 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 MSE. However, as the sample size increases, the para-

metric method has a non-decreasing bias, while the semiparametric method has decreasing bias. The variance of the semiparametric estimator $\widehat{\beta}_1$ is still larger than the parametric one. However, for $\widehat{\beta}_2$, the parametric method gives a noticeably larger bias and leads to a larger MSE than the semiparametric method. The same observations can also be made from Table 2, which reports empirical squared bias, variance, and MSE of parametric and semiparametric estimators for a panel size of 5000 and refreshment sample size of 2500.

In the bivariate normal setting, our proposed parametric and semiparametric methods perform better than CMR. When the joint distribution is correctly specified, the parametric method outperforms the other two methods. When the distribution is misspecified, however, there will be bias in the parametric estimator, while the semiparametric estimator, being free of distributional assumptions, gives a consistent performance in the presence of non-normal populations.

4.2 Effect of Weighting

Weight assignments, as shown in section 3.2, allow us to prioritize the comparison of the density function estimates over different regions of interest. To investigate the impact of weights, we generate data from three distinct distributions: bivariate normal as in subsection 4.1, a uniform distribution, and a Beta distribution. For uniform and Beta distributions, 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 as well as a minimum of $-2\sqrt{5}$ and a maximum of $2\sqrt{5}$. respectively. In all three distributions, the two-wave data have the same marginal mean of 0 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 $\alpha-$th sample quantiles for Y_1 and Y_2. The first set $e_{1,i1}$, $e_{1,i2}$ imposes no weighting, and the second set $e_{2,i1}$, $e_{2,i2}$ only considers the middle 90% of the data.

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

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). ESEs are calculated through 1000 simulation replications and serve as a benchmark for comparison, but are not available in practice. ASEs are based on the asymptotic variance in Theorem 3, which requires knowledge of true parameter values and population density functions, making it often impractical. Thus, we propose using 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 5000 and refreshment size 2500 are drawn from a bivariate normal population with 0 marginal means, variances 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 null hypotheses among the 200 replications is calculated as the empirical power for each method, which 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 gives the power functions based on the BSE (solid), the ASE (dash), and the 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 1 as the true value of β_i moves away from the hypothesized value of 0 for all three methods, which indicates that the proposed Wald test works reasonably well. More importantly, the power functions based on 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 also constructed based on both 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 BSE have empirical coverage probabilities closer to the nominal level of 95%. In contrast, the confidence intervals based on 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 (MPN) since 2013, a multiple-wave longitudinal study aimed at understanding changes in travel behavior over time. Detailed information can be found in Hoogendoorn-Lanser et al. (2015).

The MPN samples households as survey units and collects travel information by distributing questionnaires to members in each household. The MPN has initial and second wave data from 2013 and 2014, respectively. The database consists of three components: household data, personal data, and individual travel diary data. Based on the household data, there were 3572 households in the initial wave and 4685 households in the second wave. In the first wave, 2380 households provided household information and travel diary data. Among those 2380 completed cases, 1685 households continued to report their travel behaviors during the second wave of data collection, while the remaining 695 households did not respond. A refreshment sample of 1382 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 are for the complete set, incomplete set, and refreshment sample, respectively.

The Netherlands Mobility Panel is used to study travel behavior over time. Various studies have analyzed NMP data to gain insights into travel behaviors. For instance, Kroesen (2016) investigated the relationship between attitudes and travel behaviors using the complete dataset. Hoogendoorn (2015) estimated nonresponse bias by modeling nonresponse behavior with logistic regression and a MAR assumption. LaPaix (2016) explored the effects of non-random attrition on mobility rates using trip diary data. Their analysis assumed MAR assumption

and attrition was evaluated only through observed demographic data.

The above literature assumed MCAR or MAR in their analyses of the Netherlands Mobility Panel. In contrast, our study relaxes the missing mechanism assumption and considers 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 with a natural log transformation.

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 in red, while the one based on the full panel Y_1 is in green. In the right panel, the estimated marginal density of Y_2 based on the complete set is in red, and the density based on the refreshment sample is in green. The estimated marginal densities of Y_1 and Y_2 based on the complete set (in red) can be biased due to missingness in the data. In contrast, the full panel Y_1 and the refreshment sample provide more accurate estimates (in green) 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 ith subject with $W_i = 1$ if Y_2 is observed for the subject i and $W_i = 0$ otherwise. We assume an additive logistic model for the probability of $W_i = 1$ as $\pi_{\text{MNAR}} = P(W_i = 1 \mid y_1, y_2, \beta) = \text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)$. It reduces to MCAR with $\beta_1 = \beta_2 = 0$ and MAR with $\beta_2 = 0$.

Table 5 gives the estimation results of the attrition parameters and their 95% confidence intervals under three different missing mechanisms. Under MNAR, the confidence intervals for both β_1 and β_2 do not contain 0, indicating strong evidence that the missingness is related to Y_1 and Y_2. Therefore, neither MCAR nor MAR are adequate assumptions for the Netherlands Mobility Panel. In addition, the positive estimate of β_1 indicates that the probability of being observed in the second wave increases as the value of total travel time in the first wave increases, while 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. It is also consistent with the observation in Figure 4 that the complete set (red) has a density leaning toward lower values of Y_2 compared with the marginal density from the refreshment sample (green). Under MNAR, the 95% confidence intervals are constructed through bootstrapping. Figure 5 plots the sampling distributions of the bootstrapped semiparametric estimators. The red vertical lines

represent the point estimates from the original data.

6. Discussion

We extend Hirano et al. (2001)'s method for identifying and estimating the non-ignorable attrition mechanism for binary responses to continuous responses using the 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 real data analysis. 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 due to the curse of dimensionality in multivariate nonparametric density estimation. However, our method can be extended to a more flexible attrition model 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 requires further research.

Supplementary Material

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

Acknowledgments

Research reported in this publication 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 fifth International Symposium on Multivariate Analysis*, vol. 5, 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. and Wilson, D. (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. *Tech. rep.*, University of Washington.

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.

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.

— (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.

— (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.

	Obs	Y_1	Y_2	W
Complete Set	1	Y_{11}	Y_{12}	W_1=1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_c	Y_{n_c1}	Y_{n_c2}	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^r_{12}	
	$\vdots$		$\vdots$	
	n		Y^r_{n2}	

Table 1: Two wave data with refreshment sample

	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

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 5000 and refreshment sample size of 2500.

Distribution	n_1	n_2		Squared Bias (10^{-3})		Variance (10^{-3})		MSE (10^{-3})	
				$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
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

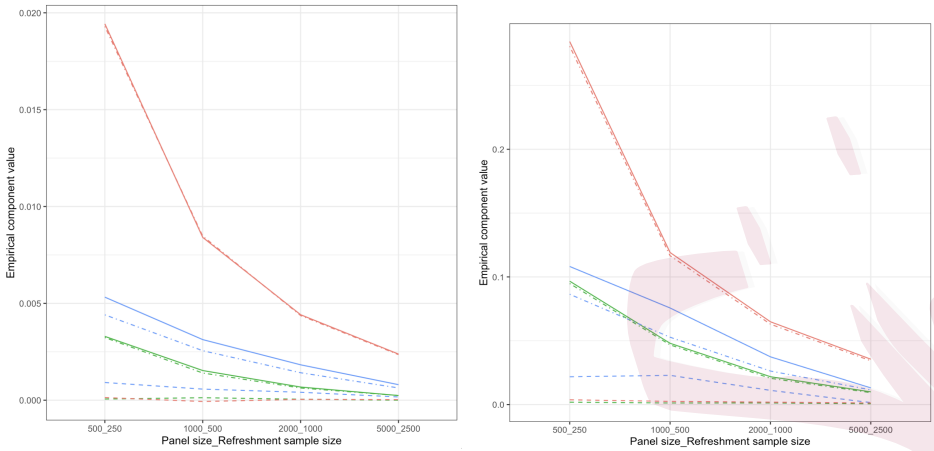
Table 3: Empirical squared bias, variance, and MSE of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for semiparametric methods with two weights e_1: no weight, e_2: a weight that uses only 90% of the data under different distributions.

		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

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 5000, and the refreshment sample size is 2500. The coverage probabilities are calculated for different true values of attrition parameters β_1 and β_2.

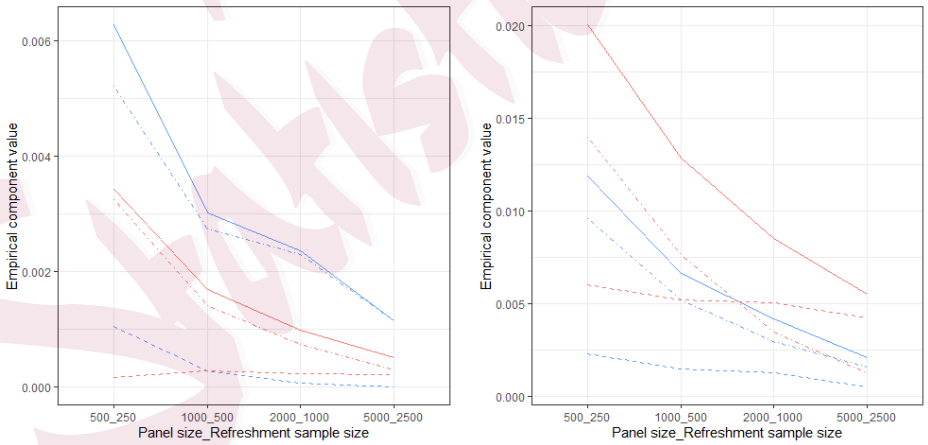
Attrition model	MCAR		MAR		MNAR	
$logit(\pi) =$	β_0		$\beta_0 + \beta_1 y_1$		$\beta_0 + \beta_1 y_1 + \beta_2 y_2$	
$\widehat{\beta}_0$	0.89	(0.80, 0.97)	0.03	(-0.47, 0.54)	7.11	(5.09, 8.91)
$\widehat{\beta}_1$			0.15	(0.06, 0.24)	0.71	(0.50, 0.88)
$\widehat{\beta}_2$					-1.64	(-1.97, -1.25)

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



(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 (green), semiparametric (blue), and CMR (red) for bivariate normal responses. Dashed, dotted-dash and solid lines represent empirical squared bias, variance, and MSE, respectively.



(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. The parametric method is plotted in red, and the semiparametric method is in cyan. For both methods, dash, dotted dash, and solid lines represent empirical squared bias, variance, and MSE, respectively.

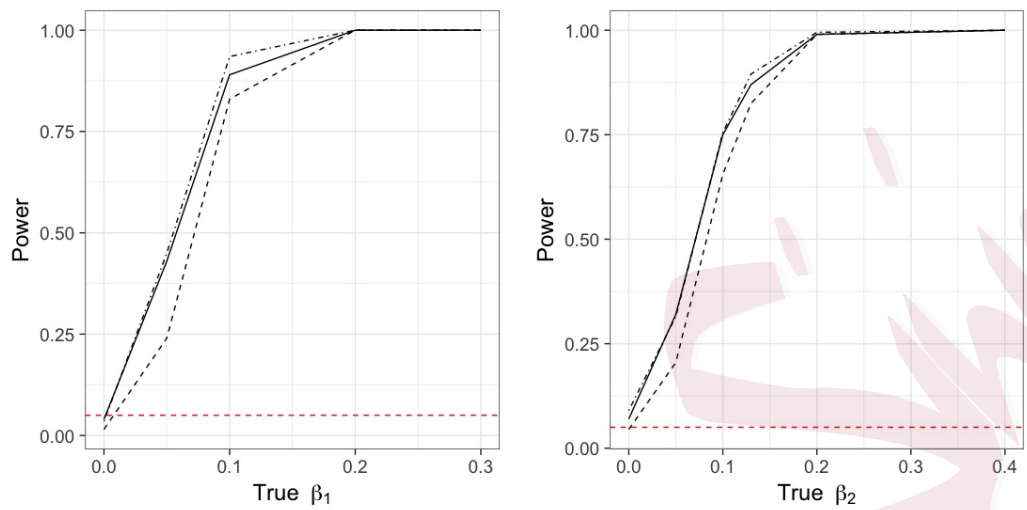


Figure 3: Power function comparison. The solid, dash and dot-dash lines represent the power functions based on the bootstrap SE (BSE), the asymptotic SE (ASE), and the empirical SE (ESE), respectively. The red dash line at the bottom locates the significance level, 0.05.

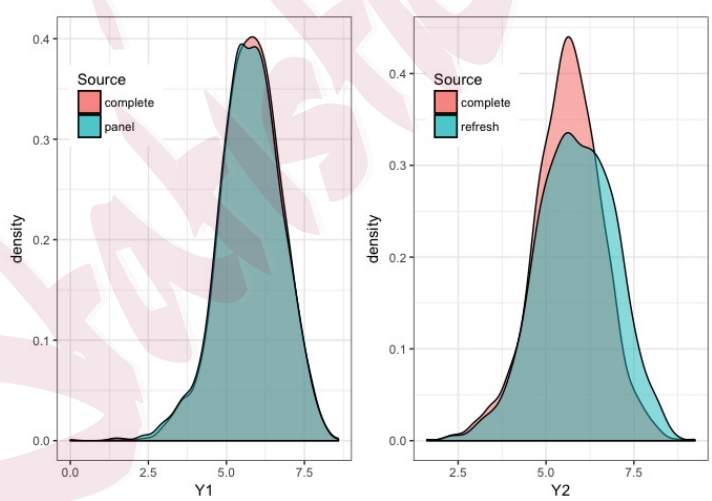


Figure 4: Marginal density comparison of MPN on the first and second wave

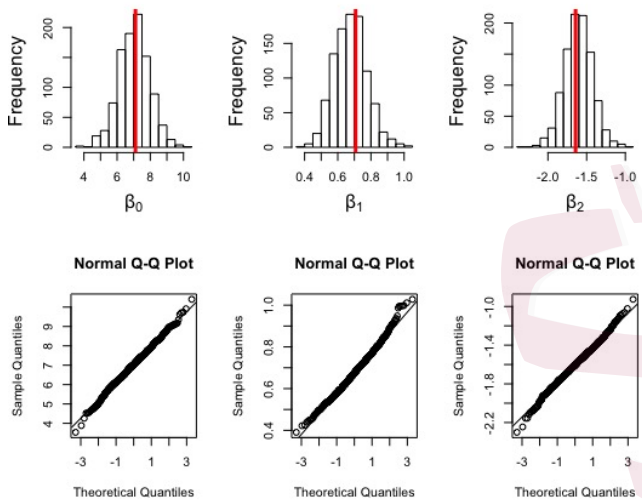


Figure 5: Sampling distributions of bootstrapped semiparametric estimators in MPN application.