7-6-2012

An efficient method of estimation for longitudinal surveys with monotone missing data

Ming Zhou
Iowa State University

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

CHAPTER 5. AN EFFICIENT METHOD OF ESTIMATION FOR LONGITUDINAL SURVEYS WITH MONOTONE MISSING DATA

A paper to be submitted to *Biometrika*

Ming Zhou and Jae Kwang Kim

Department of Statistics & Statistical Laboratory

Iowa State University

Ames, IA, 50011, USA

Abstract

Panel attrition is frequently encountered in panel sample surveys. When the panel attrition is related with the observed study variable observed in the previous years, the classical approach of nonresponse adjustment using a covariate-dependent dropout mechanism can be biased. We consider an efficient method of estimation with monotone panel attrition when the response probability depends on the previous values of study variable as well as other covariates. The proposed estimator is asymptotically optimal in the sense that it minimizes the asymptotic variance of a class of estimators that can be written as a linear combination of the unbiased estimators of the panel estimates for each wave. The proposed estimator incorporates all available information using the idea of generalized method of moments. Variance estimation is discussed and results from two limited simulation studies are also presented.

Key Words: Ignorable Missing; Generalized Method of Moments; Panel attrition; Survey in Time, Survey Sampling.

5.1 Introduction

Longitudinal surveys or panel surveys, surveys in which similar measurements are made on the same sample at different points in time, are very popular in the study of social and physical dynamics that cannot be inferred from cross-sectional surveys. Missing data in the response variable is a serious impediment to performing a valid statistical analysis in longitudinal surveys, and estimation with longitudinal missing data is quite challenging. Bollinger and David (1997, 2001), for example, used the Survey of Income and Program Participation (SIPP) data to show that estimates of food-stamp participation adjusted for nonresponse are significantly different from estimates that fail to account for nonresponse. Wun et al. (2007) and Hawkes and Plewis (2006) showed empirically that modeling the response probability is important for reducing the bias in the estimates.

A popular practice for nonresponse adjustment for panel survey assumes that the implicit response mechanism is the covariate-dependent missing, as termed by Little (1995), where the response probability depends on the base-year covariate X_i that does not change over time, but not on the study variable Y_{it} that may vary over time. The nonresponse mechanism is called ignorable if the true response probability depends only on the observed data and does not depend on unobserved random variables. In a panel survey, ignorable response mechanism means that the response probability at time t may depend on X_i and Y_{is} with $s < t$, but not on Y_{it}. The covariate-dependent missing mechanism can be quite restrictive because the dropout mechanism may not be fully explained by demographic base year covariates. For example, Korinek et al. (2007) analyzed the Current Population Survey (CPS) data using an area level model of response rate on average income and found that the response probability is strongly correlated with household income. They went on further concluding that the current adjustment method, which is essentially based on the covariate-dependent missing assumption, should be rejected.

Robins et al. (1995) have developed a method for the estimation of longitudinal regression models under ignorable nonresponse. Robins et al. (1995) assumed a working outcome regression model for Y_{it}, as well as the response propensity model for the response probability, in developing their estimator. Rotnitzky et al. (1998) extended the method of Robins et al. (1995) to nonignorable missingness. However, the Robins et al. (1995) method does not make full use of available information.

Under the response model of Robins et al. (1995), we consider an alternative method of parameter estimation that uses all available information in the longitudinal data. The basic idea for combining all available information is based on the generalized method of moments (GMM) estimation procedure of Hansen (1982). For T=1, the proposed estimator reduces to the optimal estimator considered in Cao et al. (2009). Thus, the proposed estimator is a generalized version of the existing optimal estimator for longitudinal surveys. By the orthogonal construction of the control variates, the computation of the optimal estimator is simplified.

In Section 5.2, basic setup is introduced. In Section 5.3, optimal propensity score estimator is motivated under the GMM framework. In Section 5.4, the optimal estimator under the panel survey setup is proposed. In Section 5.5, the proposed method is extended to complex survey sampling. In Section 5.6, results from two limited simulation studies are presented and concluding remarks are made in Section 5.7.

5.2 Basic Setup

Let $Y_{it} (i = 1, \ldots, n, t = 0, \ldots, T)$ be the outcome of interest measured on the ith subject at year t, X_i be the corresponding auxiliary information that is always observed and remains constant throughout different years. We use r_{it} to denote the indicator of response for subject i at year t: $r_{it} = 1$ if Y_{it} is observed and $r_{it} = 0$ otherwise. We shall regard $(X_i, r_{i0}, r_{i1}, \ldots, r_{iT}, Y_{i0}, \ldots Y_{iT})$, $i = 1, \ldots, n$, as independent and identically distributed random vectors. Assume that the baseline information for subject i, (X_i, Y_{i0}), is always observed. Our goal is to estimate $\mu_t = E(Y_{it})$, the mean of Y_{it}, for $t = 1, \ldots, T$. Denote $L_{i,t} = (X_i', Y_{i0}, Y_{i1}, \ldots, Y_{i,t})'$ be the observed values of (X, Y) for unit i up to time t. For any random variable Δ, we use $\tilde{E}$ to denote the sample average of Δ, that is

$$\tilde{E}\{\Delta\} = n^{-1} \sum_{i=1}^{n} \Delta_i. \tag{5.1}$$

If the sample is obtained from a complex sampling design, $\tilde{E}\{\Delta\}$ represent a design-unbiased estimator of $E\{\Delta\}$ based on the theory of Horvitz and Thompson (1952).

Throughout this paper, we shall assume that the missing pattern is monotone, that is,

$$r_{i,j} = 0 \Rightarrow r_{i,j+1} = 0, \forall j = 1, \ldots, T-1. \tag{5.2}$$

Although the constraint (5.2) can be somewhat restrictive, we believe that the monotone missing will cover most realistic situations for the panel attrition. The extension to non-monotone missing pattern is beyond the scope of this paper. We shall assume the following missing data mechanism:

$$Pr(r_{it} = 1 | r_{i,t-1} = 1, L_{i,T}) = Pr(r_{i,t} = 1 | r_{i,t-1} = 1, L_{i,t-1}). \tag{5.3}$$

Equation (5.3) means that the data are "missing at random" in the sense of Rubin (1976). See also Little (1995) for its meaning under longitudinal survey setup. Equation (5.3) means that, at any year t, the probability that Y_{it} is missing only depends on what is observed by time $t-1$. In other words, among subjects observed at time $t-1$, the nonresponse probability at time t is unrelated to the current and future outcomes $Y_{it}, \ldots, Y_{iT}$. The missing data mechanism in (5.3) is more realistic than the covariate-dependent missing mechanism, which is often assumed in the usual nonresponse adjustment methods that use the demographic variables as the covariates for the nonresponse model. In addition to (5.3), we assume that

$$p_{it} := P(r_{it} = 1 | r_{i,t-1} = 1, L_{i,t-1}) > \sigma > 0, \, t = 1, \ldots, T, \tag{5.4}$$

so that for each subject i, the probability of remaining in the study is bounded away from zero, which is needed to guarantee the existence of $n^{1/2}$-consistent estimators of μ_t (Robins et al., 1994). The probability p_{it} is the conditional probability of response at time t given the unit i response at time $t-1$. Assumptions (5.2) and (5.3) imply that

$$\pi_{it} := P(r_{it} = 1 | L_{i,T}) = P(r_{it} = 1 | L_{i,t-1}) = P(r_{i0} = 1) \prod_{j=1}^{t} p_{ij}. \tag{5.5}$$

The response probability π_{it} is also often called propensity score (Rosenbaum and Rubin, 1983). Denote $\pi_{i0} = P(r_{i0} = 0)$, which is assumed to be known, such as in sample surveys, the selection is under control of the investigator. The probability π_{it} is different from p_{it} in that π_{it} refers to, for subject i, the marginal probability of response at time t, while p_{it} refers to the conditional probability of a response at time t given that unit i responds at time $t-1$. Very often π_{it} depends on $L_{i,t-1}$ and the average of the observed Y_{it}'s,

$$\hat{\mu}_{t,naive} = \frac{\sum_i r_{it} Y_{it}}{\sum_i r_{it}}, \tag{5.6}$$

will in general be inconsistent for μ_t. In this case, a commonly adopted approach is to model the response probability and use the estimated response probability to obtain the propensity score adjusted (PS) estimators. We discuss the PS estimator in more detail in the upcoming section.

5.3 Optimal PS Estimation

5.3.1 PS Estimation

For simplicity, let us start from the $T=1$ case. We now absorb Y_{i0} into X_i, and denote it as X_i solely. The outcome of interest is then Y_{i1} and we are interested in estimating $\mu_1 = E(Y_{i1})$. Let the true response probability be parametrically modeled by $\pi_{i1} = \pi_1(X_i;\phi_1)$, for some function $\pi_1(.)$ known up to ϕ_1. If the maximum likelihood estimator of ϕ_1, the solution to

$$S_1(\phi_1) = \tilde{E}\left\{(r_1 - \pi_1)\frac{\partial \pi_1/\partial \phi_1}{\pi_1(1-\pi_1)}\right\} = 0, \tag{5.7}$$

denoted by $\hat{\phi}_1$, is available, then the propensity score adjusted (PS) estimator of μ_{Y_1}, denoted by $\hat{\mu}_{Y_1,PS}$, can be computed by solving

$$\hat{U}_{1,PS} := \tilde{E}\left\{\frac{r_1}{\hat{\pi}_1}(Y_1 - \mu_1)\right\} = 0, \tag{5.8}$$

for μ_1. Inverse probability weighted estimating equations have been previously considered by Horvitz and Thompson (1952), Manski and Lerman (1977), Flanders and Greenland (1991), Robins et al. (1995) among others. Strictly speaking, the PS estimator in (5.8) is also a function of $\hat{\phi}_1$. To discuss the asymptotic variance of the PS estimator, we introduce the following lemma.

Lemma 5.1. *Suppose $z_1, \ldots, z_n$ are independent and identically distributed random vectors and $\hat{\gamma}$ is the solution to $\tilde{E}\{U(z;\gamma)\} = 0$. Let $U_i(\gamma) = U(z_i;\gamma)$. If (i) $E\{U(\gamma^*)\} = 0$ and $\hat{\gamma} = \gamma^* + o_p(1)$, where γ^* is an interior point of the parameter space; (ii) $Var\{U(\gamma^*)\}$ is finite; (iii) $U(\gamma)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of γ^*; (iv) $E\{\partial U(\gamma^*)/\partial\gamma\}$ exists and is nonsingular; (v) $E\left\{\sup_{\gamma\in\mathcal{N}} \|\partial U(\gamma)/\partial\gamma\|\right\} < \infty$. Then,*

$$\hat{\gamma} - \gamma^* = -\left[E\left\{\frac{\partial U(\gamma^*)}{\partial\gamma}\right\}\right]^{-1} \tilde{E}\{U(\gamma^*)\} + o_p(n^{-1/2}). \tag{5.9}$$

Proof. This lemma is an immediate application of Lemma 2.4 and Theorem 3.1 of Newey and McFadden (1994). □

Remark 5.1. Consider $U_i(\gamma) = (\theta' - g_i(\phi)', \psi_i(\phi)')'$, where $\gamma = (\theta', \phi')'$, $S(\phi) = \sum_i \psi_i(\phi)$ is the score function for ϕ. Let $\hat{\gamma}$ be the solution to $\tilde{E}\{U(\gamma)\} = 0$, that is, $\hat{\gamma} = (\hat{\theta}', \hat{\phi}')'$, where $\hat{\theta} = \tilde{E}\{g(\hat{\phi})\}$. Assume that conditions in Lemma 5.1 are satisfied. Then (5.9) reduces to

$$\begin{pmatrix} \hat{\theta} - \theta^* \\ \hat{\phi} - \phi^* \end{pmatrix} = -\begin{bmatrix} I & -E\{\partial g(\phi^*)/\partial\phi\} \\ 0 & E\{\partial\psi(\phi^*)/\partial\phi\} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{E}\{g(\phi^*)\} \\ \tilde{E}\{\psi(\phi^*)\} \end{bmatrix} + o_p(n^{-1/2}). \tag{5.10}$$

By differentiating $E\{g(\phi)\}$ with respect to ϕ under the integral sign and evaluating at ϕ^*, we obtain $-E\{\partial g(\phi^*)/\partial\phi\} = E\{g(\phi^*)\psi(\phi^*)\} = Cov\{g(\phi^*), \psi(\phi^*)\}$, similarly $-E\{\partial\psi(\phi^*)/\partial\phi\} = E\{\psi(\phi^*)\psi(\phi^*)'\} = Var\{\psi(\phi^*)\}$ (Pierce, 1982) . Therefore, $\hat{\theta}$ can be expressed as

$$\hat{\theta} - \theta^* = \tilde{E}\{g(\phi^*)\} - Cov\{g(\phi^*), \psi(\phi^*)\}[Var\{\psi(\phi^*)\}]^{-1}\tilde{E}\{\psi(\phi^*)\} + o_p(n^{-1/2}). \tag{5.11}$$

This implies that

$$Var[\tilde{E}\{g(\hat{\phi})\}] = Var[\tilde{E}\{g(\phi^*)\}|S^\perp] + o(n^{-1}), \tag{5.12}$$

where

$$\begin{aligned} Var\{\tilde{E}\{g(\hat{\phi})|S^\perp\} := {} & Var[\tilde{E}\{g(\phi^*)\}] \\ & - Cov[\tilde{E}\{g(\phi^*)\}, S(\phi^*)][Var\{S(\phi^*)\}]^{-1}Cov[S(\phi^*), \tilde{E}\{g(\phi^*)\}]. \end{aligned} \tag{5.13}$$

Note that by (5.12),

$$Var\left\{\tilde{E}\left(\frac{r_1 Y_1}{\hat{\pi}_1}\right)\right\} \approx Var\left\{\tilde{E}\left(\frac{r_1 Y_1}{\pi_1}\right)\middle| S_1^\perp\right\} \leq Var\left\{\tilde{E}\left(\frac{r_1 Y_1}{\pi_1}\right)\right\}.$$

Such contradictory phenomena has been discussed by Rosenbaum (1987), Robins et al. (1994), Little and Vartivarian (2005), Kim and Kim (2007), Beaumont and Bocci (2009). See also Henmi and Eguchi (2004).

5.3.2 Optimal PS Estimation

We now discuss optimal PS estimation. We assume that the propensity score is computed as in (5.7). In general, the PS estimator $\hat{\mu}_{X,PS}$ applied to $\mu_X = E(X)$ is not equal to the complete sample estimator $\hat{\mu}_{X,n} = \tilde{E}\{X\}$. Thus, the complete sample estimator $\bar{X}_n$ can be used to improve the efficiency of the PS estimator. To combine all the available information, we consider minimizing the following objective function

$$Q = \begin{pmatrix} \hat{X}_1 - \mu_X \\ \hat{X}_2 - \mu_X \\ \hat{Y}_1 - \mu_1 \end{pmatrix}' Var\left\{ \begin{pmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \hat{Y}_1 \end{pmatrix} \right\}^{-1} \begin{pmatrix} \hat{X}_1 - \mu_X \\ \hat{X}_2 - \mu_X \\ \hat{Y}_1 - \mu_1 \end{pmatrix}, \tag{5.14}$$

with respect to μ_X and μ_1, where $\hat{X}_1$ and $\hat{X}_2$ are two unbiased estimators of μ_X and $\hat{Y}_1$ is an unbiased estimator of μ_1. The estimator obtained from the minimization of Q in (5.14) is often called the GMM, termed by Hansen (1982), and is very popular in econometrics. The GMM setup provides a useful tool for combining information from different sources. Under the missing data setup where X_i is always observed and Y_{i1} is subject to missingness, if we know π_{i1}, then we can evaluate $\hat{X}_1 = \tilde{E}\{X\}, \hat{X}_2 = \tilde{E}\{r_1 X/\pi_1\}, \hat{Y}_1 = \tilde{E}\{r_1 Y_1/\pi_1\}$. In this case, the optimal estimator that minimizes (5.14) is obtained by

$$\hat{\mu}_1 = \tilde{E}\left\{ r_1 Y_1/\pi_1 - (r_1/\pi_1 - 1)X'B^* \right\},$$

where

$$B^* = E\left[\tilde{E}\{(1/\pi_1 - 1)XX'\}\right]^{-1} E\left[\tilde{E}\{(1/\pi_1 - 1)XY_1\}\right].$$

In practice, we can estimate B^* by the plug-in estimator

$$\hat{\mu}_{1,opt} = \tilde{E}\left\{ r_1 Y_1/\pi_1 - (r_1/\pi_1 - 1)X'\hat{B}^* \right\}, \tag{5.15}$$

where

$$\hat{B}^* = \tilde{E}\left[\frac{r_1}{\pi_1}(1/\pi_1 - 1)XX'\right]^{-1} \tilde{E}\left[\frac{r_1}{\pi_1}(1/\pi_1 - 1)XY_1\right]$$

The estimator (5.15) is an (asymptotically) optimal estimator among the class of linear unbiased estimators. A PS estimator is said to have external consistency if it is equal to the full sample estimator $\tilde{E}\{Y_1\}$ when $Y_{i1} = X_i'\beta$ for some β for all i. Note that the optimal estimator (5.15) satisfies a calibration constraint in the sense that, if $Y_{i1} = X_i'\beta$ for some β for all i, then $\hat{B}^* = \beta$, and $\hat{\mu}_{1,opt} = \tilde{E}\{r_1(Y_1 - Y_1)/\pi_1 + Y_1\} = \tilde{E}(Y_1)$. Thus, the direct PS estimator $\hat{Y}_{1,PS}$ is not externally consistent but the optimal estimator in (5.15) is.

If the true propensity scores are unknown, we will use $\hat{X}_1 = \tilde{E}\{X\}$, $\hat{X}_2 = \hat{X}_{PS} = \tilde{E}\{r_1X/\hat{\pi}_1\}$, $\hat{Y}_1 = \hat{Y}_{1,PS} = \tilde{E}\{r_1Y_1/\hat{\pi}_1\}$, where $\hat{\pi}_1 = \pi_1(X;\hat{\phi}_1)$, with $\hat{\phi}_1$ being the maximum likelihood estimator given by (5.7). In this case, the optimal estimator of μ_X is still equal to $\tilde{E}\{X\}$, but the optimal estimator of $\hat{\mu}_{1,opt}$ in (5.15) using $\hat{\pi}_{i1}$ instead of π_{i1} is not really optimal because the covariance between $\hat{Y}_{1,PS}$ and $(\hat{X}_{PS}, \hat{X}_n)$ is different from the covariance between $\tilde{E}\{r_1Y_1/\pi_1\}$ and $(\tilde{E}\{r_1X/\pi_1\}, \hat{X}_n)$. To construct an optimal estimator, we can consider an estimator of the form

$$\hat{\mu}_{1,B} = \hat{Y}_{1,PS} - (\hat{X}_{PS} - \hat{X}_n)B$$

indexed by B and find B^* that minimizes the variance of $\hat{\mu}_{1,B}$. The solution is

$$B^* = \left\{Var(\hat{X}_{PS} - \hat{X}_n)\right\}^{-1} Cov(\hat{X}_{PS} - \hat{X}_n, \hat{Y}_{1,PS}).$$

By (5.12), we can approximate B^* by

$$B^* = \left[Var\{\tilde{E}(r_1X/\pi_1) - \hat{X}_n | S_1^{\perp}\}\right]^{-1} Cov\{\tilde{E}(r_1X/\pi_1) - \hat{X}_n, \tilde{E}(r_1Y_1/\pi_1) | S_1^{\perp}\}. \tag{5.16}$$

Thus, the optimal estimator in (5.14) with $\hat{X}_1 = \hat{X}_n, \hat{X}_2 = \hat{X}_{PS}, \hat{Y}_1 = \hat{Y}_{1,PS}$ can be obtained by minimizing

$$Q = (\hat{Z} - \mu_Z)'\{Var(\hat{Z}_0 | S_1^{\perp})\}^{-1}(\hat{Z} - \mu_Z), \tag{5.17}$$

where $\hat{Z} = (\hat{X}_n, \hat{X}_{PS}, \hat{Y}_{1,PS})'$, $\hat{Z}_0 = (\hat{X}_n, \tilde{E}\{r_1X/\pi_1\}, \tilde{E}\{r_1Y_1/\pi_1\})$ and $\mu_Z = (\mu_X, \mu_X, \mu_{Y_1})$. The optimal Q term in (5.17) can be also obtained by minimizing the augmented Q term given by

$$Q^* = \begin{pmatrix} \hat{Z} - \mu_Z \\ S_1 \end{pmatrix}' \left\{ Var \begin{pmatrix} \hat{Z}_0 \\ S_1 \end{pmatrix} \right\}^{-1} \begin{pmatrix} \hat{Z} - \mu_Z \\ S_1 \end{pmatrix}, \tag{5.18}$$

where $S_1(\phi_1)$ is the score function of ϕ_1, defined in (5.7). To see this, note that Q^* can be decomposed into

$$\begin{aligned} Q^* &= \{\hat{Z} - \mu_Z - Var(\hat{Z}_0)^{-1}Cov(\hat{Z}_0, S_1)S_1\}'\{Var(\hat{Z}_0|S_1^{\perp})\}^{-1} \\ &\quad \times \{\hat{Z} - \mu_Z - Var(\hat{Z}_0)^{-1}Cov(\hat{Z}_0, S_1)S_1\} + S_1'Var(S_1)^{-1}S_1 \\ &= Q + S_1'Var(S_1)^{-1}S_1, \end{aligned}$$

where Q is defined in (5.17). Because $S_1'Var(S_1)^{-1}S_1$ does not involve μ_Z, the optimal estimator of μ_Z can be also be computed by minimizing Q^* term in (5.18). Thus, the effect of using the estimated propensity score can be easily taken into account by simply adding the score function for ϕ_1 into the Q term. Furthermore, as shall be discussed in Theorem 5.2, the inclusion of the score function into the GMM makes the linearization for variance estimation easy. The following example gives an explicit formula for the optimal PS estimator when ϕ_1 is estimated by its maximum likelihood estimator.

Example 5.1. Under the response model where the score function for ϕ_1 is

$$S_1(\phi_1) = \tilde{E}\left\{\left(\frac{r_1}{\pi_1(\phi_1)} - 1\right)h(\phi_1)\right\},$$

The coefficient B^* corresponding to the optimal PS estimators in the family

$$\hat{\mu}_{1,B} = \hat{Y}_{1,PS} - (\hat{X}_{PS} - \hat{X}_n)'B,$$

is given by

$$B^* = \{V_{XX} - V_{XS}V_{SS}^{-1}V_{SX}\}^{-1}\{V_{XY} - V_{XS}V_{SS}^{-1}V_{SY}\},$$

where

$$\begin{pmatrix} V_{XX} & V_{XY} & V_{XS} \\ V_{YX} & V_{YY} & V_{YS} \\ V_{SX} & V_{SY} & V_{SS} \end{pmatrix} = Var\left[\begin{pmatrix} \tilde{E}\{(r_1/\pi_1 - 1)X\} \\ \tilde{E}\{r_1Y_1/\pi_1\} \\ \tilde{E}\{(r_1/\pi_1 - 1)h\} \end{pmatrix}\right].$$

Thus, a consistent estimator of B^* is given by

$$\hat{B}^* = (I_p, O)\left[\tilde{E}\left\{\frac{r_1}{\hat{\pi}_1}\left(\frac{1}{\hat{\pi}_1} - 1\right)\begin{pmatrix} X \\ h(X) \end{pmatrix}\begin{pmatrix} X \\ h(X) \end{pmatrix}'\right\}\right]^{-1}\tilde{E}\left\{\frac{r_1}{\hat{\pi}_1}\left(\frac{1}{\hat{\pi}_1} - 1\right)\begin{pmatrix} X \\ h(X) \end{pmatrix}Y_1\right\}, \tag{5.19}$$

where $p = dim(X)$, and the resulting optimal estimator is

$$\hat{Y}_{1,opt} = \hat{Y}_{1,PS} - (\hat{X}_{PS} - \hat{X}_n)'\hat{B}^*. \tag{5.20}$$

The estimator in (5.20) is equal to the optimal estimator presented in Cao et al. (2009) in the context of a doubly robust estimator. Similar but slightly different approach was proposed by Tan (2006). However, our derivation of the optimal estimator in (5.20) is different from those of Cao et al. (2009) and Tan (2006). In addition, the GMM setup used in deriving (5.20) can be easily generalized in longitudinal missing data, which will be discussed in the following section.

5.4 Proposed Method for Longitudinal Missing

The proposed optimal estimator in Section 5.3 is based on the GMM setup and it can be easily extended to the problem of optimal estimation with longitudinal missing. To correctly account for the ignorable dropout mechanism in (5.3), we shall assume a parametric model for p_{it}, given by $p_{it}(L_{i-1};\phi_t)$. Note that we do not make any explicit assumptions for the marginal distribution of $L_{i,T}$, we only use the response model, which is attractive in handling unit nonresponse in sample surveys. The partial likelihood regarding ϕ_t's is then

$$L(\phi_1, \ldots, \phi_T) = \prod_{i=1}^{n}\prod_{t=1}^{T}\left[p_{it}^{r_{i,t}}(1-p_{it})^{1-r_{i,t}}\right]^{r_{i,t-1}}. \tag{5.21}$$

The corresponding score function is then

$$\bar{S}_T := (S_1(\phi_1), \ldots, S_T(\phi_T))' = \left(\frac{\partial \log L(\phi_1, \ldots, \phi_T)}{\partial \phi_1}, \ldots, \frac{\partial \log L(\phi_1, \ldots, \phi_T)}{\partial \phi_T}\right), \tag{5.22}$$

where

$$S_t(\phi_t) = n\tilde{E}\left\{r_{t-1}(r_t - p_t)\frac{\partial p_t/\partial \phi_t}{p_t(1-p_t)}\right\} \tag{5.23}$$

is the score function associated with the conditional response probability. A commonly adopted parametric model for p_t is the logistic regression model

$$p_t = \frac{1}{1+\exp\{-\phi_t' L_{t-1}\}}. \tag{5.24}$$

Under this logistic regression model (5.24), the score function in (5.23) reduces to

$$S_t(\phi_t) = n\tilde{E}\left\{r_{t-1}(r_t - p_t)L_{t-1}\right\}. \tag{5.25}$$

At each year t, we can obtain PS estimators for $\mu_X, \mu_1, \ldots, \mu_t$, using $\hat{\pi}_t$. Thus, we have $T+1$ estimators of μ_X and $T-t+1$ estimators of μ_t computed by the inverse probability weights at years $t, \ldots, T$. To illustrate it, denote operator $\mathcal{M}_t$ as

$$\mathcal{M}_t(\Delta) = \tilde{E}\left\{\frac{r_t\Delta}{\hat{\pi}_t}\right\}, \; t = 0, 1, \ldots. \tag{5.26}$$

Then we can obtain a collection of PS estimators for L_T, as shown in Table 5.1. In year $t = 2$, for example, $\mathcal{M}_2(X)$ is available for μ_X, $M_2(Y_1)$ is available for μ_1, and $\mathcal{M}_2(Y_2)$ is available for μ_2.

Table 5.1 List of available unbiased estimators of the parameters for each year

Parameters	$t=0$	$t=1$	$t=2$	$t=3$	$\cdots$	$t=T$
μ_X	$\mathcal{M}_0(X)$	$\mathcal{M}_1(X)$	$\mathcal{M}_2(X)$	$\mathcal{M}_3(X)$	$\cdots$	$\mathcal{M}_T(X)$
μ_1		$\mathcal{M}_1(Y_1)$	$\mathcal{M}_2(Y_1)$	$\mathcal{M}_3(Y_1)$	$\cdots$	$\mathcal{M}_T(Y_1)$
μ_2			$\mathcal{M}_2(Y_2)$	$\mathcal{M}_3(Y_2)$	$\cdots$	$\mathcal{M}_T(Y_2)$
μ_3				$\mathcal{M}_3(Y_3)$	$\cdots$	$\mathcal{M}_T(Y_3)$
$\vdots$					$\ddots$	$\vdots$
μ_T						$\mathcal{M}_T(Y_T)$

To incorporate all available information, we use adopt GMM method in Section 5.3. Denote $u_t = r_t/p_t - 1$, $t = 1, \ldots, T$, and

$$\psi_{t-1} = \begin{pmatrix} r_0u_1p_1L_0 \\ r_1u_2p_2L_1 \\ \vdots \\ r_{t-1}u_tp_tL_{t-1} \end{pmatrix}, \tag{5.27}$$

where $L_0 = (X', Y_0)'$ and $L_j = (X', Y_0, Y_1, \ldots, Y_j)'$ for $j = 1, \ldots, t-1$. Note that at each year t, $E(r_{t-1}u_tp_tL_{t-1}) = 0$, because $E(r_{t-1}u_tp_tL_{t-1}|L_{t-1}, r_1 = \ldots = r_{t-1} = 1) = r_{t-1}p_tL_{t-1}E(u_t|L_{t-1}, r_1 = \ldots r_{t-1} = 1) = 0$. Thus, $E\{\psi_{t-1}\} = 0$. At each year t, we are able to construct PS estimators for $\mu_X, \mu_{Y_1}, \ldots, \mu_t$ using $\tilde{E}(r_t X/\pi_t)$, $\tilde{E}(r_t Y_1/\pi_t)$, …, $\tilde{E}\{r_t Y_t/\pi_t\}$ respectively. Similar to the $t = 1$ case, at year t, we get the following control variates

$$\xi_{t-1} = \begin{pmatrix} \frac{r_0}{\pi_0} u_1 L_0 \\ \frac{r_1}{\pi_1} u_2 L_1 \\ \vdots \\ \frac{r_{t-1}}{\pi_{t-1}} u_t L_{t-1} \end{pmatrix}. \tag{5.28}$$

By the definition of the response probabilities, we have $r_{t-1}u_t/\pi_{t-1} = r_t/\pi_t - r_{t-1}/\pi_{t-1}$ and $E\{\xi_{t-1}\} = 0$. Therefore, we propose the following optimal estimator for $E(Y_t)$ as the minimizer to the following quadratic Q_t, with respect to μ_t, using the fact that $E\{\psi_{t-1}\} = 0, E\{\xi_{t-1}\} = 0$,

$$Q_t = \begin{pmatrix} \tilde{E}\{r_t Y_t/\hat{\pi}_t\} - \mu_t \\ \tilde{E}\{\hat{\xi}_{t-1}\} \\ \tilde{E}\{\psi_{t-1}\} \end{pmatrix}' \left\{ \hat{V}ar \begin{pmatrix} \tilde{E}\{r_t Y_t/\pi_t\} \\ \tilde{E}\{\xi_{t-1}\} \\ \tilde{E}\{\psi_{t-1}\} \end{pmatrix} \right\}^{-1} \begin{pmatrix} \tilde{E}\{r_t Y_t/\hat{\pi}_t\} - \mu_t \\ \tilde{E}\{\hat{\xi}_{t-1}\} \\ \tilde{E}\{\psi_{t-1}\} \end{pmatrix}, \tag{5.29}$$

where $\hat{\xi}_{t-1}$ is ξ_{t-1} after plugging in the maximum likelihood estimator $\hat{\phi}_1, \ldots, \hat{\phi}_t$ given by (5.22). The control variate $\hat{\xi}_{t-1}$ is included to incorporate all available information up to year $t-1$ and the control variate $\tilde{E}(\psi_{t-1})$ is included to incorporate the score equation for $(\phi_1', \ldots, \phi_{t-1}')'$. For $t = 1$, $\tilde{E}(\hat{\xi}_0) = \hat{X}_{PS} - \hat{X}_n$ and $\tilde{E}(\psi_{t-1}) = n^{-1}S_1$. Note that we can write $\bar{S}_T = n\tilde{E}\{\psi_{T-1}\}$. For example, when $T = 3$,

$$\xi_2 = \begin{pmatrix} \frac{r_0}{\pi_0} u_1 L_0 \\ \frac{r_1}{\pi_1} u_2 L_1 \\ \frac{r_2}{\pi_2} u_3 L_2 \end{pmatrix}, \ \psi_2 = \begin{pmatrix} (r_1 - p_1 r_0) L_0 \\ (r_2 - p_2 r_1) L_1 \\ (r_3 - p_2 r_2) L_2 \end{pmatrix}.$$

Intuitively speaking, when estimating $E(Y_3)$, we have four PS estimators for X, which are $\tilde{E}(X)$, $\tilde{E}(r_1 X/\hat{\pi}_1)$, $\tilde{E}(r_2 X/\hat{\pi}_2)$, $\tilde{E}(r_3 X/\hat{\pi}_3)$; three PS estimators for Y_1, i.e., $\tilde{E}(r_1 Y_1/\hat{\pi}_1)$, $\tilde{E}(r_2 Y_2/\hat{\pi}_2)$, $\tilde{E}(r_3 Y_1/\hat{\pi}_3)$; two PS estimators for Y_2, $\tilde{E}(r_2 Y_2/\hat{\pi}_2), \tilde{E}(r_3 Y_2/\hat{\pi}_2)$. Those nine PS estimators produce six atomic control variates represented by $\hat{\xi}_2$, in the sense that any difference between two PS estimators for estimating the same mean, say $\tilde{E}(r_j z/\hat{\pi}_j) - \tilde{E}(r_i z/\hat{\pi}_i)$, can be written as a linear combination of $\tilde{E}\{\hat{\xi}_2\}$, where z can be any past information before year t. Formally speaking, the following theorem gives our optimal PS estimator for μ_t for $t = 1, \ldots, T$. Note that, because of the orthogonality of $r_0 u_1, \ldots, r_{t-1} u_t$, the t subvectors of $\tilde{E}(\xi_{t-1})$ and $\tilde{E}(\psi_{t-1})$

are also orthogonal and $\hat{V}\{\tilde{E}(\xi_{t-1})\}$ and $\hat{V}\{\tilde{E}(\psi_{t-1})\}$ are block diagonal matrices. This orthogonality of the control variates makes the computation of the resulting optimal estimator simple.

Theorem 5.1. *Under the regularity conditions given in Appendix 5.A.1 and the response model (5.24) such that the score equation for $(\phi_1', \ldots, \phi_T')'$ is $\tilde{E}(\psi_{T-1}) = 0$, for each year t, the coefficient B_t^* corresponding to the optimal estimator of $\mu_t = E\{Y_t\}$ among the class*

$$\tilde{E}(r_t Y_t/\hat{\pi}_t) - B_t' \tilde{E}(\hat{\xi}_{t-1}),$$

is given by $B_t^* = ({B_{1t}^*}', \ldots, {B_{tt}^*}')'$, where

$$\begin{aligned} B_{jt}^* = (I_{dim(L_{j-1})}, O) E^{-1} &\left\{ r_{j-1} \left(\frac{1}{p_j} - 1 \right) \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1} \end{pmatrix} \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1} \end{pmatrix}' \right\} \\ \times E &\left\{ \left(\frac{1}{p_j} - 1 \right) \begin{pmatrix} \frac{1}{\pi_{i-1}} L_{i-1} \\ p_i L_{i-1} \end{pmatrix} \frac{r_t Y_t}{\pi_t} \right\}. \end{aligned} \tag{5.30}$$

A consistent estimator for B_{jt}^* is

$$\begin{aligned} \hat{B}_{j,t} = (I_{dim(L_{j-1})}, O) \tilde{E}^{-1} &\left\{ \frac{r_t \hat{\pi}_{j-1}}{\hat{\pi}_t} \left(\frac{1}{\hat{p}_j} - 1 \right) \begin{pmatrix} \frac{1}{\hat{\pi}_{j-1}} L_{j-1} \\ \hat{p}_j L_{j-1} \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{\pi}_{j-1}} L_{j-1} \\ \hat{p}_j L_{j-1} \end{pmatrix}' \right\} \\ \times \tilde{E} &\left\{ \left(\frac{1}{\hat{p}_j} - 1 \right) \begin{pmatrix} \frac{1}{\hat{\pi}_{i-1}} L_{i-1} \\ \hat{p}_i L_{i-1} \end{pmatrix} \frac{r_t Y_t}{\hat{\pi}_t} \right\}. \end{aligned} \tag{5.31}$$

The resulting optimal estimator that minimizes (5.29) is

$$\hat{Y}_{t,opt} = \tilde{E}\{r_t Y_t/\hat{\pi}_t\} - \sum_{j=1}^{t} \hat{B}_{j,t}' \tilde{E} \left\{ \frac{r_{j-1} \hat{u}_j L_{j-1}}{\hat{\pi}_{j-1}} \right\}, \tag{5.32}$$

where $\hat{u}_j = r_j/\hat{p}_j - 1$ and $\hat{p}_j = p_j(L_{j-1}; \hat{\phi}_j)$.

Proof. See Appendix. □

Remark 5.2. For $t = 1, r_0 \equiv 1, \pi_0 \equiv 1$, the estimator is

$$\tilde{E}\{\hat{r}_1 Y_1/\hat{\pi}_1\} - \hat{B}_{1,1}' \tilde{E} \left\{ \frac{r_0}{\pi_0} \hat{u}_1 L_0 \right\} = \tilde{E} \left\{ \frac{r_1}{\hat{p}_1} Y_1 - \hat{B}_{1,1}' \left(\frac{r_1}{\hat{p}_1} - 1 \right) X \right\},$$

where

$$\hat{B}_{1,1} = (I, O)\left[\tilde{E}\left\{\frac{r_1}{\hat{p}_1}\left(\frac{1}{\hat{p}_1}-1\right)\begin{pmatrix} X \\ \hat{p}_1 X\end{pmatrix}\begin{pmatrix} X \\ \hat{p}_1 X\end{pmatrix}'\right\}\right]^{-1}\tilde{E}\left\{\frac{r_1}{\hat{p}_1}\left(\frac{1}{\hat{p}_1}-1\right)\begin{pmatrix} X \\ \hat{p}_1 X\end{pmatrix} Y_1\right\},$$

which is the same estimator as given in Example 5.1.

We now discuss variance estimation of the optimal estimator in (5.32). Strictly speaking, $\hat{Y}_{t,opt}$ is a function of $(\hat{\phi}_1, \ldots, \hat{\phi}_t)$ and should then be written as $\hat{Y}_{t,opt}(\hat{\phi}_1, \ldots, \hat{\phi}_t)$. We show in Theorem 5.2 that we can safely ignore the effects of $\hat{\phi}_1, \ldots, \hat{\phi}_t$ in $\hat{Y}_{t,opt}$ for linearization variance estimation. That is, $\hat{Y}_{t,opt}(\hat{\phi}_1, \ldots, \hat{\phi}_t) = \hat{Y}_{t,opt}(\phi_1^*, \ldots, \phi_t^*) + o_p(n^{-1/2})$, which is often referred to as Randles (1982) condition. See Kim and Rao (2009), for details.

Theorem 5.2. *Under the regularity conditions in Appendix 5.A.1, $\hat{Y}_{t,opt}$ in (5.32) is asymptotically linear with influence function η_t, where*

$$\eta_t = \frac{r_t Y_t}{\pi_t} - \sum_{j=1}^{t} D'_{j,t} r_{t-1} u_t \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}\end{pmatrix}, \tag{5.33}$$

where

$$D_{j,t} = E^{-1}\left\{r_{j-1}\left(\frac{1}{p_j}-1\right)\begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}\end{pmatrix}\begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}\end{pmatrix}'\right\} E\left\{\left(\frac{1}{p_j}-1\right)\begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}\end{pmatrix}\frac{r_t Y_t}{\pi_t}\right\}.$$

Thus,

$$\sqrt{n}(\hat{Y}_{t,opt} - \mu_t) \xrightarrow{d} N\{0, Var(\eta_t)\} \tag{5.34}$$

and also

$$\hat{V}^{-1/2}(\hat{Y}_{t,opt} - \mu_t) \xrightarrow{d} N(0, 1), \tag{5.35}$$

where

$$\hat{V} = n^{-1}(n-1)^{-1}\tilde{E}\{\hat{\eta}_t - \tilde{E}(\hat{\eta}_t)\}^2, \tag{5.36}$$

and $\hat{\eta}_t$ is η_t with the estimated parameters plugged-in.

Proof. See Appendix. □

Remark 5.3. We obtain $\hat{\mu}_{t,opt}$ by minimizing Q_t in (5.29) with respect to μ_t. One may consider estimating $\mu_1, \ldots, \mu_T$ simultaneously by minimizing the following term

$$\tilde{Q}_T = \begin{pmatrix} \tilde{E}\{X\} - \mu_X \\ \tilde{E}\{r_1 Y_1/\hat{\pi}_1\} - \mu_1 \\ \vdots \\ \tilde{E}\{r_T Y_T/\hat{\pi}_T\} - \mu_T \\ \tilde{E}\{\hat{\xi}_{T-1}\} - E\{\xi_{T-1}\} \\ \tilde{E}\{\psi_{T-1}\} - E\{\psi_{T-1}\} \end{pmatrix}' \left\{ \hat{V}ar \begin{pmatrix} \tilde{E}\{X\} \\ \tilde{E}\{r_1 Y_1/\pi_1\} \\ \vdots \\ \tilde{E}\{r_T Y_t/\pi_T\} \\ \tilde{E}\{\xi_{T-1}\} \\ \tilde{E}\{\psi_{T-1}\} \end{pmatrix} \right\}^{-1} \begin{pmatrix} \tilde{E}\{X\} - \mu_X \\ \tilde{E}\{r_1 Y_1/\hat{\pi}_1\} - \mu_1 \\ \vdots \\ \tilde{E}\{r_T Y_T/\hat{\pi}_T\} - \mu_T \\ \tilde{E}\{\hat{\xi}_{T-1}\} - E\{\xi_{T-1}\} \\ \tilde{E}\{\psi_{T-1}\} - E\{\psi_{T-1}\} \end{pmatrix}, \tag{5.37}$$

with respect to $(\mu_X', \mu_1, \ldots, \mu_T)'$. It can be shown that under monotone missing pattern, minimizing $\tilde{Q}_T$ to estimate $\mu_1, \ldots, \mu_T$ simultaneously is equivalent to minimizing $\tilde{Q}_T$ in (5.37) for each μ_t (see Appendix). The dimension of the vector in (5.37) is $2qT + T^2 + 1$, while the dimension associated with Q_t in (5.29) is $2qt + t^2 - t + 1$, where $q = dim(X)$.

5.5 Extension to Complex Survey Sampling

In this section, we extend the result to complex survey sampling by considering a finite population indexed by $U_N = \{1, 2, \ldots, N\}$ with known population size N. Let $\mathcal{F}_N = \{(X_i', Y_{i1}, \ldots, Y_{iT})' \mid i = 1, \ldots, N\}$. At each time t, Y_{it} is subject to missingness indicated by r_{it}, which takes the value 1 if unit i is responding and takes the value 0 otherwise. We shall assume monotone missing pattern as described in (5.2), and adopt missing at random mechanism as in (5.3). Let A denote the set of indices for the subjects in a sample selected by a probability sampling, with fixed sample size n and design weights $\omega_i, i = 1, \ldots, N$. Assume that the sampling indicators $I\{i \in A\}, i = 1, \ldots, N$, are independent of missing indicators r_{it}. We use notations $\tilde{E}$, $\tilde{E}_A$ defined as

$$\tilde{E}\{\Delta\} = N^{-1} \sum_{i=1}^{N} \Delta_i, \qquad \tilde{E}_A\{\Delta\} = N^{-1} \sum_{i \in A} \omega_i \Delta_i. \tag{5.38}$$

The parameters of interest are the population means of the study variables at different time points,

$$\mu_t = \tilde{E}\{Y_t\} = N^{-1}\sum_{i=1}^{N} Y_{it}, \quad t = 1, \ldots, T. \tag{5.39}$$

Under logistic regression model in (5.24), the score function for estimating ϕ_t is

$$S_t(\phi_t) = \tilde{E}_A\{r_{t-1}(r_t - p_t)L_{t-1}/w\} = \sum_{i\in A} r_{i,t-1}(r_{i,t} - p_{i,t})L_{i,t-1}. \tag{5.40}$$

The PS estimator for μ_t in (5.39) then is

$$\hat{Y}_{t,PS} = \tilde{E}_A\left(\frac{r_t Y_t}{\hat{\pi}_t}\right) = \frac{1}{N}\sum_{i\in A}\omega_i \frac{r_{it}Y_{it}}{\hat{\pi}_{it}}. \tag{5.41}$$

To apply the GMM methodology, we shall adopt ξ_{t-1} in (5.28), ψ_{t-1} in (5.27), and construct a Q_t term similar to (5.29). Note that $E\{\tilde{E}_A(r_tY_t/\pi_t)|\mathcal{F}_N\} = \tilde{E}(r_tY_t/\pi_t) = \mu_t$. Since $E[I\{i \in A\}r_{i,t-1}(r_{it}/p_{it} - 1)|\mathcal{F}_N] = E[I\{i \in A\}|\mathcal{F}_N] \cdot E\{r_{i,t-1}(r_{i,t} - p_{it})|\mathcal{F}_N\} = 0$, we have $E\{E_A(\xi_{t-1})|\mathcal{F}_N\} = 0$, and $E\{\tilde{E}_A(\psi_{t-1}/w)|\mathcal{F}_N\} = 0$. Thus we can consider the Q_t term similar to (5.29) as

$$Q_t = \begin{pmatrix} \tilde{E}_A\{r_tY_t/\hat{\pi}_t\} - \mu_t \\ \tilde{E}_A\{\hat{\xi}_{t-1}\} \\ \tilde{E}_A\{\psi_{t-1}/w\} \end{pmatrix}' \hat{V}ar^{-1}\left\{\begin{pmatrix} \tilde{E}_A\{r_tY_t/\pi_t\} \\ \tilde{E}_A\{\xi_{t-1}\} \\ \tilde{E}_A\{\psi_{t-1}/w\} \end{pmatrix}\Bigg|\mathcal{F}_N\right\}\begin{pmatrix} \tilde{E}_A\{r_tY_t/\hat{\pi}_t\} - \mu_t \\ \tilde{E}_A\{\hat{\xi}_{t-1}\} \\ \tilde{E}_A\{\psi_{t-1}/w\} \end{pmatrix}. \tag{5.42}$$

The details of the key steps for deriving the optimal solution to minimize Q_t in (5.42) are given in the Appendix. To discuss the asymptotic properties of the PS estimators in the complex survey, the following conditions are assumed in addition to the regularity conditions (C1)-(C6) stated for Theorem 5.1.

(C7) The design weight is bounded from above and below, that is,

$$0 < K_l \leq nN^{-1}\omega_i \leq K_u < \infty,$$

for all $i = 1, \ldots, N$, uniformly in n, where K_l and K_u are fixed constants.

(C8) The sample moments with design weight converges to the population moments, that is,

$$\frac{1}{N}\sum_{i\in A}\omega_i \mathbf{u}_i\mathbf{u}_i' = \frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i' + o_p(1),$$

for any $\mathbf{u}_i$ with finite second moments.

Corollary 5.1. *Let $\mathcal{F}_N = \{(X_i', Y_{i,1}, \ldots, Y_{i,T})' \mid i = 1, \ldots N\}$ be a finite population subject to missingness at $t = 1, \ldots, T$. A sample of size n is selected using design weights ω_i. Subject to conditions (C1) - (C8), under monotone missing pattern and response model in (5.24) such that the score equation for $(\phi_1', \ldots, \phi_T')'$ is $N\tilde{E}\{\psi_{T-1}\} = 0$, the optimal estimator of μ_t among the class $\tilde{E}_A\{r_t Y_t/\hat{\pi}_t\} - B_t'\tilde{E}_A\{\hat{\psi}_{t-1}\}$ is $\tilde{E}_A\{r_t Y_t/\hat{\pi}_t\} - B_t^{*\prime}\tilde{E}_A\{\hat{\psi}_{t-1}\}$, where $B_t^* = (B_{1t}^{*}{}', \ldots, B_{tt}^{*}{}')'$ and*

$$\begin{aligned} B_{jt}^* = (I_{dim(L_{j-1})}, 0) & \left[\tilde{E} \left\{ w\pi_{j-1} \left(\frac{1}{p_j} - 1 \right) \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}/w \end{pmatrix} \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}/w \end{pmatrix}' \right\} \right]^{-1} \\ & \times \tilde{E} \left\{ w \left(\frac{1}{p_j} - 1 \right) \begin{pmatrix} \frac{1}{\pi_{j-1}} L_{j-1} \\ p_j L_{j-1}/w \end{pmatrix} Y_t \right\}, \end{aligned} \tag{5.43}$$

which can be consistently estimated by

$$\begin{aligned} \hat{B}_{j,t} = (I_{dim(L_{j-1})}, 0) & \left[\tilde{E}_A \left\{ w\frac{r_t}{\hat{\pi}_t}\hat{\pi}_{j-1} \left(\frac{1}{\hat{p}_j} - 1 \right) \begin{pmatrix} \frac{1}{\hat{\pi}_{j-1}} L_{j-1} \\ \hat{p}_j L_{j-1}/w \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{\pi}_{j-1}} L_{j-1} \\ \hat{p}_j L_{j-1}/w \end{pmatrix}' \right\} \right]^{-1} \\ & \times \tilde{E}_A \left\{ w\frac{r_t}{\hat{\pi}_t} \left(\frac{1}{\hat{p}_j} - 1 \right) \begin{pmatrix} \frac{1}{\hat{\pi}_{j-1}} L_{j-1} \\ \hat{p}_j L_{j-1}/w \end{pmatrix} Y_t \right\}, \end{aligned} \tag{5.44}$$

The resulting optimal estimator for minimizing (5.42) is

$$\hat{Y}_{t,opt} = \tilde{E}_A\{r_t Y_t/\hat{\pi}_t\} - \sum_{j=1}^{t} \hat{B}_{j,t}'\tilde{E}_A\left\{\frac{r_{j-1}\hat{u}_j L_{j-1}}{\hat{\pi}_{j-1}}\right\}, \tag{5.45}$$

where $\hat{u}_{ij} = r_{ij}/\hat{p}_{ij} - 1$, $\hat{\pi}_{ij} = \prod_{k=1}^{j}\hat{p}_{ik}$ and $\hat{p}_{ij} = p_j(L_{i,j-1};\hat{\phi}_j)$.

Remark 5.4. When $t = 1$, note that we assume no missing in the baseline year, i.e. $\pi_0 = 1$, the optimal estimator for $N^{-1}\sum_{i=1}^{N} Y_{i1}$ is

$$\begin{aligned} \hat{Y}_{1,opt} &= \tilde{E}_A\{r_1 Y_1/\hat{\pi}_1\} - \hat{B}_{11}'\tilde{E}_A\{(r_1/\hat{\pi}_1 - 1)X\} \\ &= N^{-1}\sum_{i\in A}\omega_i\frac{r_{i1}Y_{i1}}{\hat{\pi}_{i1}} - N^{-1}\sum_{i\in A}\omega_i\left(\frac{r_{i1}}{\hat{\pi}_{i1}} - 1\right)X'\hat{B}_{11}, \end{aligned}$$

where $\hat{B}_{11}$ is

$$(I, O)\left\{\sum_{i\in A}\omega_i^2\frac{r_{i1}}{\hat{\pi}_{i1}}\left(\frac{1}{\hat{p}_{i1}} - 1\right)\begin{pmatrix} X_i \\ \frac{\hat{p}_{i1}X_i}{w_i}\end{pmatrix}\begin{pmatrix} X_i \\ \frac{\hat{p}_{i1}X_i}{w_i}\end{pmatrix}'\right\}^{-1} \times \left\{\sum_{i\in A}\omega_i^2\frac{r_{i1}}{\hat{\pi}_{i1}}\left(\frac{1}{\hat{p}_{i1}} - 1\right)\begin{pmatrix} X_i \\ \frac{\hat{p}_{i1}X_i}{w_i}\end{pmatrix} Y_{i1}\right\}.$$

Denote

$$\eta_{i,t} = \frac{r_{it}Y_{it}}{\pi_{it}} - \sum_{j=1}^{t} B_{j,t}^{*}{}' r_{i,j-1}u_{i,j}\frac{L_{i,j-1}}{\pi_{i,j-1}} - \sum_{j=1}^{t} C_{j,t}^{*}{}' r_{i,j-1}u_{i,j}\frac{p_{i,j}L_{i,j-1}}{w_i}, \tag{5.46}$$

where $(B_{j,t}^{*}{}', C_{j,t}^{*}{}') = D_{j,t}^{*}{}'$, and

$$\begin{aligned} D_{jt}^{*} = & \left[\tilde{E}\left\{ w\pi_{j-1}\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}/w\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}/w\end{pmatrix}'\right\}\right]^{-1} \\ & \times \tilde{E}\left\{ w\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}/w\end{pmatrix} Y_t\right\}. \end{aligned} \tag{5.47}$$

A consistent estimator of $D_{j,t}$ is

$$\begin{aligned} \hat{D}_{j,t} = & \left[\tilde{E}_A\left\{ w\frac{r_t}{\hat{\pi}_t}\hat{\pi}_{j-1}\left(\frac{1}{\hat{p}_j}-1\right)\begin{pmatrix}\frac{1}{\hat{\pi}_{j-1}}L_{j-1}\\ \hat{p}_jL_{j-1}/w\end{pmatrix}\begin{pmatrix}\frac{1}{\hat{\pi}_{j-1}}L_{j-1}\\ \hat{p}_jL_{j-1}/w\end{pmatrix}'\right\}\right]^{-1} \\ & \times \tilde{E}_A\left\{ w\frac{r_t}{\hat{\pi}_t}\left(\frac{1}{\hat{p}_j}-1\right)\begin{pmatrix}\frac{1}{\hat{\pi}_{j-1}}L_{j-1}\\ \hat{p}_jL_{j-1}/w\end{pmatrix} Y_t\right\}. \end{aligned} \tag{5.48}$$

Let $\hat{\eta}_{i,t}$ be the corresponding estimator of $\eta_{i,t}$ in (5.46) with $\hat{D}_{j,t}, \hat{\pi}_{i,j}, \hat{p}_{i,j}$, then $\hat{Y}_{t,opt}$ in (5.45) can be written as

$$\hat{Y}_{t,opt} = \frac{1}{N}\sum_{i\in A}\omega_i\hat{\eta}_{i,t} = \frac{1}{n}\sum_{i\in A}\frac{n\omega_i}{N}\hat{\eta}_{i,t}.$$

By similar arguments in the proof of Theorem 5.2,

$$\tilde{E}_A(\hat{\eta}_t) = \tilde{E}_A(\eta_t) + o_p(n^{-1/2}),$$

and we can apply the standard complete sample method to estimate the variance of $\tilde{E}_A(\eta_t)$, which is asymptotically equivalent to the variance of $\tilde{E}_A(\hat{\eta}_t)$ (see Kim and Rao, 2009).

To calculate $Var\{\tilde{E}_A(\eta_t)|\mathcal{F}_N\}$, the *reverse framework* of Fay (1992), Shao and Steel (1999), Kim and Rao (2009) is used. Specifically, denote $\mathbf{r}_t = \{r_{11}, \ldots, r_{Nt}\}$ and $\bar{\mathbf{r}}_t = \{\mathbf{r}_1, \ldots, \mathbf{r}_t\}$. Then

$$Var\{\tilde{E}_A(\eta_t)|\mathcal{F}_N\} = V_1 + V_2 = E[Var\{\tilde{E}_A(\eta_t)|\bar{\mathbf{r}}_t, \mathcal{F}_N\}|\mathcal{F}_N] + Var[E\{\tilde{E}_A(\eta_t)|\bar{\mathbf{r}}_t, \mathcal{F}_N\}|\mathcal{F}_N]. \tag{5.49}$$

For any g with finite second moment, we assume that $N^{-1}\sum_{i\in A}\sum_{i\in A}\Omega_{ij}g_ig_j$ is a design unbiased estimator of $Var\{\tilde{E}(g)|\mathcal{F}_N\}$, where Ω_{ij} depends on the joint inclusion probability. Then $Var\{\tilde{E}_A(\eta_t)|\bar{\mathbf{r}}_t,\mathcal{F}_N\}$ in (5.49) can be estimated by

$$\hat{V}_1(\eta) = N^{-2}\sum_{i\in A}\sum_{j\in A}\Omega_{ij}\eta_{i,t}\eta_{j,t}.$$

To show the consistency of $\hat{V}_1$ for V_1 in (5.49), we assume that finite fourth moments exist for variables stated in (C4),

$$\sum_{i=1}^{N}|\Omega_{N.ij}| = O(n^{-1}N), \tag{5.50}$$

and

$$Var[nVar\{\tilde{E}_A(\eta_t)|\bar{\mathbf{r}}_t,\mathcal{F}_N\}|\mathcal{F}_N] = o_p(1).$$

Consequently, $\hat{V}_1(\eta)$ is consistent for V_1 and $\hat{V}_1(\hat{\eta})$ is also consistent for V_1 under same conditions (see Kim et al., 2006). The second term V_2 in (5.49) is $V_2 = Var[E\{\tilde{E}_A(\eta_t)|\bar{\mathbf{r}}_t,\mathcal{F}_N\}|\mathcal{F}_N] = Var\{\tilde{E}(\eta_t)|\mathcal{F}_N\}$. Note that $E(\eta_t|\mathcal{F}_N) = E(r_tY_t/\pi_t|\mathcal{F}_N) + 0 = Y_t$, $r_{j-1}u_j/\pi_{j-1} = r_j/\pi_j - r_{j-1}/\pi_{j-1}$, then

$$\begin{aligned}
\eta_t - E(\eta_t|\mathcal{F}_N) &= \left(\frac{r_t}{\pi_t}-1\right)Y_t - \sum_{j=1}^{t} r_{j-1}u_j{D^*_{j,t}}'\begin{pmatrix} L_{j-1}/\pi_{j-1} \\ L_{j-1}p_j/w\end{pmatrix} \\
&= \sum_{j=1}^{t}\left(\frac{r_j}{\pi_j}-\frac{r_{j-1}}{\pi_{j-1}}\right)Y_t - \sum_{j=1}^{t} r_{j-1}u_j{D^*_{j,t}}'\begin{pmatrix} L_{j-1}/\pi_{j-1} \\ L_{j-1}p_j/w\end{pmatrix} \\
&= \sum_{j=1}^{t} r_{j-1}u_j\left\{\frac{Y_t}{\pi_{j-1}} - \sum_{j=1}^{t}{D^*_{j,t}}'\begin{pmatrix} L_{j-1}/\pi_{j-1} \\ L_{j-1}p_j/w\end{pmatrix}\right\}.
\end{aligned}$$

Recall that $E(r_{j-1}u_j|\mathcal{F}_N) = 0, j = 1,\ldots,T$ and $E(r_{i-1}u_ir_{j-1}u_j|\mathcal{F}_N) = \pi_{j-1}(1/p_j-1)I(i=j)$, for any i,j. Then, the form of V_2 is

$$V_2 = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{t}\frac{1}{\pi_{i,j-1}}\left(\frac{1}{p_{ij}}-1\right)\left\{Y_{it} - \pi_{i,j-1}\sum_{j=1}^{t}{D^*_{j,t}}'\begin{pmatrix} L_{i,j-1}/\pi_{i,j-1} \\ L_{i,j-1}p_{i,j}/w_i\end{pmatrix}\right\}^2, \tag{5.51}$$

and it can be estimated by

$$\hat{V}_2 = N^{-2}\sum_{i\in A}\omega_i\sum_{j=1}^{t}\frac{1}{\hat{\pi}_{i,j-1}}\left(\frac{1}{\hat{p}_{ij}}-1\right)\left\{Y_{it} - \hat{\pi}_{i,j-1}\sum_{j=1}^{t}\hat{D}'_{j,t}\begin{pmatrix} L_{i,j-1}/\hat{\pi}_{i,j-1} \\ L_{i,j-1}\hat{p}_{i,j}/w_i\end{pmatrix}\right\}^2. \tag{5.52}$$

Under (C8), we have $\hat{V}_2 = V_2 + o_p(N^{-1})$. Therefore, $\hat{V}\{\tilde{E}_A(\hat{\eta}_t)\} = \hat{V}_1 + \hat{V}_2$ is consistent for the variance $\hat{Y}_{t,opt}$ in (5.45).

The order of the first term V_1 is $V_1 = O_p(n^{-1})$, while the order of the second term V_2 is $V_2 = O_p(N^{-1})$. Thus, when the sampling fraction n/N is negligible, that is, $n/N = o(1)$, the second term V_2 can be ignored, and $\hat{V}_1$ would be a consistent estimator for the total variance.

5.6 Simulation Study

To test our theory and to examine the performance of the proposed estimator for finite sample sizes, we performed two simulation studies. In the first simulation study, we used a linear regression model with serial correlation. The model is

$$Y_0 = X/2 + e_0,\ Y_t = 1 + X/2 + Y_{t-1} + e_t, \text{ for } t > 1,$$

where $X \sim N(0,1)$, and e_t's are independent and identically distributed as $N(0,1)$. The missing indicator r_t follows the following distribution:

$$P(r_t = 1 | X, Y_{t-1}, r_{t-1} = 1) = \frac{1}{1 + \exp[-2.5 - X + \{Y_{t-1} - (t-1)\}/2]},$$

and there are no missing data in the baseline year. In this simulation setup, the true mean of Y_t is $E(Y_t) = t$. The parameters of interest are $\mu_t = E(Y_t)$, for $t = 1, 2, 3$. We computed five estimators for each parameter. The estimators include $\tilde{E}\{Y_t\}$, the full sample estimator under no missingness; $\tilde{E}\{r_t Y_t\}/\tilde{E}\{r_t\}$, the naive estimator using the simple mean of the responding part of the sample; $\tilde{E}\{r_t Y_t/\hat{\pi}_t\}$, the direct PS estimator; $\hat{Y}_{t,opt}$, our optimal propensity score adjusted estimator in (5.32). In addition, we considered an estimator from the class of estimators proposed by Robins et al. (1995) based on weighted estimating equations. Specifically, Let $\hat{Y}_{t,RRZ}$ be solution to

$$\tilde{E}\left[\frac{r_t}{\hat{\pi}_t}\{Y_t - \mu_t - \beta'_{1,t}(X - \tilde{E}[X])\}\begin{pmatrix} 1 \\ X - \tilde{E}(X) \end{pmatrix}\right] = 0, \tag{5.53}$$

which gives

$$\hat{Y}_{t,RRZ} = \frac{\tilde{E}\{r_t Y_t/\hat{\pi}_t\}}{\tilde{E}\{r_t/\hat{\pi}_t\}} - \hat{\beta}'_{1,t}\left(\frac{\tilde{E}\{r_t Y_t/\hat{\pi}_t\}}{\tilde{E}\{r_t/\hat{\pi}_t\}} - \bar{X}_n\right). \tag{5.54}$$

We used $B = 10,000$ Monte Carlo samples of size $n = 500$ for this simulation. The response rates for $t = 1, 2, 3$, are $0.90, 0.83, 0.76$ respectively. The simulation results in Table 5.2 show that the naive estimator is severely biased as expected, and the other three PS estimators (direct, RRZ, optimal) are all nearly unbiased. The RRZ estimator is more efficient than the direct PS estimator because the regression model approximately holds. However, the RRZ estimator is less efficient than the optimal estimator.

We also computed a variance estimator of the optimal estimator using the formula in (5.36). The relative biases of the variance estimator in (5.36), for $t = 1, 2, 3$, are $\mathbf{0.0260}$, $\mathbf{0.0197}$, $-\mathbf{0.0280}$ respectively. Thus, the simulation results show good finite sample performance of the proposed variance estimator.

Table 5.2 Comparison for different methods when $n = 500, T = 3$ with Monte Carlo sample size 10,000 for simulation study 1.

Parameter	Estimator	Bias	Var	StdMSE
μ_1	Full	-0.0004	0.0061	100
	Naive	0.0224	0.0067	120
	PS	-0.0006	0.0063	105
	RRZ	-0.0006	0.0063	105
	Optimal	-0.0006	0.0064	105
μ_2	Full	-0.0010	0.0105	100
	Naive	-0.0756	0.0121	169
	PS	-0.0025	0.0122	116
	RRZ	-0.0026	0.0116	111
	Optimal	-0.0018	0.0115	109
μ_3	Full	-0.0015	0.0161	100
	Naive	-0.3029	0.0186	687
	PS	-0.0063	0.0522	325
	RRZ	-0.0088	0.0235	147
	Optimal	-0.0022	0.0189	118

In the second simulation study, we used a nonlinear type regression model with serial correlation. The model is

$$Y_0 = X/3 + Z/3 + e_0,\ Y_t = 1 + X/3 + Z/3 + e_t, \text{ for } t > 1,$$

where $X \sim N(0, 1)$, $Z = sgn(X)\sqrt{|X|} + \epsilon$, with sgn being the sign function, ϵ and e_t's are independent and identically distributed $N(0,1)$ random variables. The missing indicator r_t follows the following distribution:

$$P(r_t = 1|X, Y_{t-1}, r_{t-1} = 1) = \frac{1}{1 + \exp[-2.5 - X + \{Y_{t-1} - (t-1)\}/2]}, \tag{5.55}$$

and there are no missing data in the baseline year. In this simulation setup, the true mean of Y_t is $E(Y_t) = t$. The parameters of interest again are $\mu_t = E(Y_t)$, for $t = 1, 2, 3$.

Here we used $B = 10,000$ Monte Carlo samples of size $n = 500$ for this simulation. The response rates for $t = 1, 2, 3$ are $0.90, 0.82, 0.74$ respectively. The simulation results in Table 5.3 show the same tendency as Table 5.2. The relative biases of the variance estimator using the formula in (5.36), for $t = 1, 2, 3$, are **0.0137**,$-$**0.0115**, $-$**0.0671** respectively. At time $t = 3$, the relative efficiency of the proposed estimator over the RRZ estimator is 167%, which is greater than 124% of the first simulation study, and it is because the working regression model assumed in the RRZ model does not hold in the sample generated by (5.55).

Table 5.3 Comparison for different methods when $n = 500, T = 3$ with Monte Carlo sample size 10,000 for simulation study 2.

Parameter	Estimator	Bias	Var	StdMSE
μ_1	Full	0.0014	0.0079	100
	Naive	0.0234	0.0089	119
	PS	0.0013	0.0083	105
	RRZ	0.0012	0.0083	105
	Optimal	0.0014	0.0083	105
μ_2	Full	0.0015	0.0149	100
	Naive	-0.1624	0.0169	291
	PS	0.0008	0.0197	132
	RRZ	0.0001	0.0175	118
	Optimal	0.0017	0.0161	108
μ_3	Full	0.0018	0.0238	100
	Naive	-0.5958	0.0274	1605
	PS	-0.0157	0.1892	795
	RRZ	-0.0188	0.0514	217
	Optimal	0.0030	0.0309	130

5.7 Conclusion

We have considered the problem of estimating population mean for longitudinal data with monotone missing patterns. The proposed method uses a parametric response model where the response probability at time t depends on the available observations at time $t-1$, that is, on $(X', Y_1, \ldots, Y_{t-1})'$. We used a logistic regression model for the response probability, but the proposed method can be easily extended to other response probability models that use an explicit parametric form for the response probability.

The proposed method makes the best use of all (asymptotically) unbiased estimators available for each wave of the panel survey. The way we combine the information is based on the GMM technique and the resulting estimator is asymptotically optimal among a class of estimators that can be written as linear combinations of the unbiased estimators of the panel estimates for each wave. The proposed method is directly applicable to the case when the baseline year sample is selected with a complex probability sample. Variance estimation using linearization method is relatively straightforward.

The proposed method requires that the missing pattern be monotone. If the proposed method is applied to non-monotone missing patterns, estimation of response probability at time t can be more complicated because $Y_{i,t-1}$ are not always observed for non-monotone missing case. Extension of the proposed method to non-monotone missing data will be an important topic for future research.

5.A Proofs and Discussions

5.A.1 Proof of Theorem 5.1

Let $h_{i,t}(\phi_t) = \partial \mathrm{logit}(p_{it})/\partial\phi_t$ where $\mathrm{logit}(p) = \log\{p/(1-p)\}$, $H_{i,t} = (\xi'_{i,t-1}, \psi'_{i,t-1})'$. Throughout the following arguments, unless explicitly pointed out, we shall suppress the notation of true parameters ϕ_t^* such that all expectations are evaluated at the true parameters. We shall assume the following regularity conditions.

(C1) The conditional response probabilities are bounded from below uniformly, that is, there exists a fixed positive constant σ such that $p_{it} > \sigma$ for $i = 1, \ldots, n; t = 1, \ldots, T$ uniformly.

(C2) The solution $\hat{\phi}_t$ to $S_t(\phi_t) = 0$ satisfies $\hat{\phi}_t = \phi_t^* + o_p(1)$ for $t = 1, \ldots, T$.

(C3) $p_{it}(\phi_t)$ is twice continuously differentiable in the neighborhood of ϕ_t^* for $t = 1, \ldots, T$.

(C4) $X, Y_t, h_t(\phi_t^*), \partial h_t(\phi_t^*)/\partial \phi_t$ have finite second moments for $t = 1, \ldots, T$.

(C5) $Var\{H_{i,T}\}$ is nonsingular, $E\{\partial H_T/\partial \bar{\phi}_T\}$ exists and is nonsingular.

(C6) There exists a neighborhood $\mathcal{N}_t$ of ϕ_t^* such that $E\{\sup_{\phi_t \in \mathcal{N}_t} \|h_t(\phi_t)\|\} < \infty$, $E\{\sup_{\phi_t \in \mathcal{N}_t} \|h_t(\phi_t) h_t(\phi_t)'\|\} < \infty$ and $E\{\sup_{\phi_t \in \mathcal{N}_t} \|\partial h_t/\partial \phi_t\|\} < \infty$ for $t = 1, \ldots, T$.

Proof. The optimal B_t^* that minimizes the variance of $\tilde{E}\{r_t Y_t/\hat{\pi}_t\} - B_t' \tilde{E}\{\hat{\xi}_{t-1}\}$ is given by

$$Var[\tilde{E}\{\hat{\xi}_{t-1}\}] B_t^* = Cov(\tilde{E}\{\hat{\xi}_{t-1}\}, \tilde{E}\{r_t Y_t/\hat{\pi}_t\}).$$

Let $U_{i,t}(\gamma) = (\mu_t - r_{it} Y_{it}/\pi_{it}, \xi_{i,t-1}', \psi_{i,t-1}')'$, where $\gamma = (\mu_t, \bar{\phi}_t')'$ with $\bar{\phi}_t = (\phi_1', \ldots, \phi_t')'$. First of all, conditions Lemma 5.1 (i) (ii) hold by (C1), (C2), (C4). For example, because $\pi_{it} = \prod_{j=1}^t p_{ij} \pi_0 \geq \sigma^t$, $|r_{i,t}/\pi_{i,t} - r_{i,t-1}/\pi_{i,t-1}| \leq 2/\sigma^t$, $E\{r_t Y_t^2/\pi_t^2\} \leq E\{Y_t^2\}/\sigma^2$, $E[\|(r_t/\pi_t - r_{t-1}/\pi_{t-1}) L_{t-1}\|^2] \leq 2E\|L_{t-1}\|^2/\sigma^2$. Also, (C3) implies (iii), (C5) implies (iv). Note that $p_{it}(1 - p_{it}) h_{i,t} = \partial p_{it}/\partial \phi_t$, thus $E\{\sup_{\phi_t \in \mathcal{N}_t} \|\partial p_{it}/\partial \phi_t\|\} \leq E\{\sup_{\phi_t \in \mathcal{N}_t} \|h_{it}(\phi_t)\|\}/4 < \infty$, $\|\partial \pi_{it}/\partial \phi_k\| = \|\partial p_{ik}/\partial \phi_k \prod_{j \neq k} p_{ij}(\phi_j)\| \leq \|\partial p_{ik}/\partial \phi_k\|$. Moreover, $\|\partial\{(r_{i,t} - r_{i,t-1} p_{it}) h_{i,t}\}/\partial \phi_t\|$ $= \| - r_{i,t-1} (\partial p_{it}/\partial \phi_t) h_{it}' + (r_{i,t} - r_{i,t-1} p_{i,t}) \partial h_{i,t}/\partial \phi_t\| \leq \|h_{i,t} h_{i,t}'\|/4 + 2\|\partial h_{i,t}/\partial \phi_t\|$. Therefore, (C6) implies (v). Note that under logistic response model in (5.24), $h_{i,t} = L_{i,t-1}$, (C6) would automatically hold. Although in the following arguments, we adopt the logistic regression model in (5.24), the derivation shall carry through without extra effort. By similar arguments in the remark of Lemma 5.1, we have

$$\begin{aligned}
\hat{\bar{\phi}}_t - \bar{\phi}_t^* &= -E\left\{\frac{\partial \bar{S}_t(\bar{\phi}_t^*)}{\partial \bar{\phi}_t}\right\}^{-1} \bar{S}_t(\bar{\phi}_t^*) + o_p(n^{-1/2}) \\
\tilde{E}\{\hat{\xi}_{t-1}\} &= \tilde{E}\{\xi_{t-1}(\bar{\phi}_t^*)\} - E\left\{\frac{\partial \xi_{t-1}(\bar{\phi}_t^*)}{\partial \bar{\phi}_t}\right\} E\left\{\frac{\partial \bar{S}_t(\bar{\phi}_t^*)}{\partial \bar{\phi}_t}\right\}^{-1} \bar{S}_t(\bar{\phi}_t^*) + o_p(n^{-1/2}) \\
\tilde{E}\{r_t Y_t \hat{\pi}_t\} &= \tilde{E}\{r_t Y_t/\pi_t(\bar{\phi}_t^*)\} - E\left\{\frac{\partial (r_t Y_t/\pi_t(\bar{\phi}_t^*))}{\partial \bar{\phi}_t}\right\} E\left\{\frac{\partial \bar{S}_t(\bar{\phi}_t^*)}{\partial \bar{\phi}_t}\right\}^{-1} \bar{S}_t(\bar{\phi}_t^*) + o_p(n^{-1/2}).
\end{aligned}$$

By similar argument in the remark, (see also Pierce, 1982), we have

$$E\left\{\frac{\partial \xi_{t-1}}{\partial \bar{\phi}_t}\right\} = -Cov(\xi_{t-1}, \bar{S}_t) = -Cov(\xi_{t-1}, \psi_{t-1}),\ E\left\{\frac{\partial \bar{S}_t(\bar{\phi}_t)}{\partial \bar{\phi}_t}\right\} = Var(\bar{S}_t) = nVar(\psi_{t-1}).$$

Therefore,

$$
\begin{aligned}
Var[\tilde{E}\{\hat{\xi}_{t-1}\}] &= Var[\tilde{E}\{\xi_{t-1}\}] - Cov(\xi_{t-1}, \bar{S}_t)Var(\bar{S}_t)^{-1}Cov(\bar{S}_t, \xi_{t-1}) + o(n^{-1}) \\
Cov(\tilde{E}\{\hat{\xi}_{t-1}\}, \tilde{E}\{r_tY_t/\hat{\pi}_t\}) &= Cov(\tilde{E}\{\xi_{t-1}\}, \tilde{E}\{r_tY_t/\pi_t\}) \\
&\quad - Cov(\xi_{t-1}, \bar{S}_t)Var(\bar{S}_t)^{-1}Cov(\bar{S}_t, r_tY_t/\pi_t) + o(n^{-1}).
\end{aligned}
$$

Let

$$
Var\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}\end{pmatrix} = E\left\{\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}\end{pmatrix}\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}\end{pmatrix}'\right\} = \begin{pmatrix}V_{LL,t} & V_{LS,t}\\ V_{LS,t} & V_{SS,t}\end{pmatrix}, \begin{pmatrix}V_{LY,t}\\ V_{SY,t}\end{pmatrix} = E\left\{\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}\end{pmatrix} r_tY_t/\pi_t\right\},
$$

then

$$
B_t^* = (V_{LL,t} - V_{LS,t}V_{SS,t}^{-1}V_{SL,t})^{-1}(V_{LY,t} - V_{LS,t}V_{SS,t}^{-1}V_{SY,t}) + o_p(1). \tag{5.56}
$$

We can write B_t^* as

$$
B_t^* = (I, O)\begin{pmatrix}V_{LL,t} & V_{LS,t}\\ V_{LS,t} & V_{SS,t}\end{pmatrix}^{-1}\begin{pmatrix}V_{LY,t}\\ V_{SY,t}\end{pmatrix} + o_p(1). \tag{5.57}
$$

Notice that $E\{r_{i-1}u_i|L_{i-1}\} = 0, E(r_{i-1}u_i^2|L_{i-1}) = r_{i-1}\left(1/p_i - 1\right), E(r_{i-1}u_ir_{j-1}u_j|L_{i-1}) = 0$ for $i < j$, we have

$$
\begin{aligned}
V_{LL,t} &= E\left[diag\left\{\left(\frac{1}{p_1} - 1\right)\frac{r_0}{\pi_0^2}L_0L_0', \cdots, \left(\frac{1}{p_t} - 1\right)\frac{r_{t-1}}{\pi_{t-1}^2}L_{t-1}L_{t-1}'\right\}\right] \\
V_{LS,t} &= E\left[diag\left\{\left(\frac{1}{p_1} - 1\right)\frac{r_0p_1}{\pi_0}L_0L_0', \cdots, \left(\frac{1}{p_t} - 1\right)\frac{r_{t-1}p_t}{\pi_{t-1}}L_{t-1}L_{t-1}'\right\}\right] \\
V_{SS,t} &= E\left[diag\left\{\left(\frac{1}{p_1} - 1\right)r_0p_1^2L_0L_0', \cdots, \left(\frac{1}{p_t} - 1\right)r_{t-1}p_t^2L_{t-1}L_{t-1}'\right\}\right] \\
V_{LY,t} &= E\left\{\left(\frac{1}{p_1} - 1\right)\frac{L_0}{\pi_0}\frac{r_tY_t}{\pi_t}, \cdots, \left(\frac{1}{p_t} - 1\right)\frac{L_{t-1}}{\pi_{t-1}}\frac{r_tY_t}{\pi_t}\right\} \\
V_{SY,t} &= E\left\{\left(\frac{1}{p_1} - 1\right)p_1L_0\frac{r_tY_t}{\pi_t}, \cdots, \left(\frac{1}{p_t} - 1\right)p_tL_{t-1}\frac{r_tY_t}{\pi_t}\right\}.
\end{aligned}
$$

All the V matrices or vectors can be written as the form of diagonal blocks. If V is a matrix, then $V = diag(V(1), \ldots, V(t))$, where $dim\{V(j)\} = dim(L_{j-1}) \times dim(L_{j-1})$. If V is a vecotr, then $V = (V(1)', \ldots, V(t)')'$, where $dim\{V(j)\} = dim(L_{j-1}) \times 1$. Then $B_t^* = ({B_{1t}^*}', \ldots, {B_{1t}^*}')'$, where

$$
\begin{aligned}
B_{j,t}^* &= \{V_{LL,t}(j) - V_{LS,t}(j)V_{SS,t}^{-1}(j)V_{SL,t}(j)\}^{-1}\{V_{LY,t}(j) - V_{LS,t}(j)V_{SS,t}(j)^{-1}V_{SY,t}(j)\} \\
&= (I_{dim(L_{j-1})}, 0)\left[E\left\{r_{j-1}\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}\end{pmatrix}'\right\}\right]^{-1} \\
&\quad \times E\left\{\frac{r_t}{\pi_t}\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1}\\ p_jL_{j-1}\end{pmatrix}Y_t\right\}.
\end{aligned}
$$

$\square$

5.A.2 Proof of Theorem 5.2

Proof.

$$\hat{Y}_{t,opt} = \tilde{E}\{r_tY_t/\hat{\pi}_t\} - \hat{E}\{\hat{\xi}_{t-1}\}'\hat{B}_t - \tilde{E}\{\hat{\psi}_{t-1}\}'\hat{C}_t.$$

Denote parameter $\gamma = (B, C, \phi)'$ and $\gamma^* = (B_t^*, C_t^*, \bar{\phi}_t^*)$, then define

$$\mu_1(\gamma) = E_\gamma^*\{r_tY_t/\pi_t\} - B'E_{\gamma^*}\{\xi_{t-1}\} - C'E_{\gamma^*}\{\psi_{t-1}\}.$$

Then under regularity conditions (C1) - (C4), we are able to do the following derivatives,

$$
\begin{aligned}
\frac{\partial\mu_1(\gamma)}{\partial B}\Big|_{\gamma=\gamma^*} &= E_{\gamma^*}\{\xi_{t-1}\}|_{\gamma=\gamma^*} = 0 \\
\frac{\partial\mu_1(\gamma)}{\partial C}\Big|_{\gamma=\gamma^*} &= E_{\gamma^*}\{\psi_{t-1}\}|_{\gamma=\gamma^*} = 0.
\end{aligned}
$$

Moreover, notice that $\bar{S}_t = n\tilde{E}\{\psi_{t-1}\}$. Under conditions (C1)-(C6), by the results we have shown in Theorem 5.1, using the same notations, we have

$$\frac{\partial\mu_1(\gamma)}{\partial\phi}\Big|_{\gamma=\gamma^*} = V_{YS,t} - B_t^{*\prime}V_{LS,t} - C_t^{*\prime}V_{SS,t}$$

To show that $\frac{\partial\mu_1(\gamma)}{\partial\phi}\Big|_{\gamma=\gamma^*} = 0$, it suffices to show that

$$V_{SS,t}^{-1}V_{YS,t} - V_{SS,t}^{-1}V_{SL,t}B_t^* = C_t^* = (O, I)\begin{pmatrix}V_{LL,t} & V_{LS,t}\\ V_{SL,t} & V_{SS,t}\end{pmatrix}^{-1}\begin{pmatrix}V_{LY,t}\\ V_{SY,t}\end{pmatrix}. \tag{5.58}$$

Note that

$$B_t^* = (I, O)\begin{pmatrix}V_{LL,t} & V_{LS,t}\\ V_{SL,t} & V_{SS,t}\end{pmatrix}^{-1}\begin{pmatrix}V_{LY,t}\\ V_{SY,t}\end{pmatrix},$$

and we can show the following equality

$$V_{SS,t}^{-1}V_{YS} = (V_{SS,t}^{-1}V_{SL,t}, I)\begin{pmatrix} V_{LL,t} & V_{LS,t} \\ V_{SL,t} & V_{SS,t} \end{pmatrix}^{-1}\begin{pmatrix} V_{LY,t} \\ V_{SY,t} \end{pmatrix},$$

(5.58) then follows. Therefore, the Randles (1982) condition is satisfied, and

$$\hat{Y}_{t,opt} = \tilde{E}(r_tY_t/\pi_t) - \sum_{j=1}^{t} D_{j,t}^{*}{}'\tilde{E}\left\{r_{j-1}u_j\begin{pmatrix} \frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1} \end{pmatrix}\right\} + o_p(n^{-1/2}), \tag{5.59}$$

where

$$D_{j,t}^{*} = E^{-1}\left\{r_{j-1}\left(\frac{1}{p_j}-1\right)\begin{pmatrix} \frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1} \end{pmatrix}\begin{pmatrix} \frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1} \end{pmatrix}'\right\}E\left\{\left(\frac{1}{p_j}-1\right)\begin{pmatrix} \frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1} \end{pmatrix}\frac{r_tY_t}{\pi_t}\right\}.$$

Let η_t be the random quantity as given in (5.33), then $\hat{Y}_{t,opt} = \tilde{E}(\eta_t) + o_p(n^{-1/2})$. Because η_t has second moment, by central limit theorem, (5.34) holds. Now we shall show that $Var(\eta_t)$ can be consistently estimated by $\hat{V} = (n-1)^{-1}\tilde{E}\{(\hat{\eta}_t - \bar{\hat{\eta}}_t)^2\}$, where

$$\hat{\eta}_{i,t} = r_{it}Y_{it}/\hat{\pi}_{it} - \sum_{j=1}^{t}\hat{D}_{j,t}'r_{i,j-1}\hat{u}_{i,j}(L_{i,j-1}'/\hat{\pi}_{i,j-1}, \hat{p}_{ij}L_{i,j-1}')'.$$

Note that we have already shown $\tilde{E}\{\hat{\eta}_t\} = \tilde{E}(\eta_t) + o_p(n^{-1/2})$, it then suffices to show that $\tilde{E}\{\hat{\eta}_t\hat{\eta}_t'\} = \tilde{E}\{\eta_t\eta_t'\} + o_p(1)$. By (C1), (C2), (C4), (C6), there exists a neighborhood $\bar{\mathcal{N}}_t$ of $\bar{\phi}_t^{*}$ such that $E[\sup_{\bar{\phi}_t\in\bar{\mathcal{N}}_t}\|\eta_t\|] < \infty, E[\sup_{\bar{\phi}_t\in\bar{\mathcal{N}}_t}\|\eta_t\eta_t'\|] < \infty$. By Lemma 4.3 of Newey and McFadden (1994), we have $\hat{D}_{j,t} = D_{j,t}^{*} + o_p(1)$ and $\tilde{E}\{\hat{\eta}_t\hat{\eta}_t'\} = \tilde{E}\{\eta_t\eta_t'\} + o_p(1)$. Therefore, $\hat{V} = (n-1)^{-1}\tilde{E}\{(\eta_t - \bar{\eta}_t)^2\} + o_p(n^{-1}) = n^{-1}Var(\eta_t) + o_p(n^{-1})$. That is $\hat{V}/\{Var(\eta_t)/n\} = 1 + o_p(1)$. □

5.A.3 Comment for Remark 5.3

The following comment shows that whether estimating $\mu_X, \mu_1, \ldots, \mu_T$ simultaneously or not does not matter using our GMM approach. Denote $\boldsymbol{\phi} = (\phi_1', \ldots, \phi_T')'$, $\theta = (\mu_X, \mu_1, \ldots, \mu_T)'$.

Let

$$g(\mathcal{X};\theta,\boldsymbol{\phi}) = \begin{pmatrix} \tilde{E}\{r_0 X/\pi_0\} - \mu_X \\ \tilde{E}\{r_1 Y_1/\pi_1\} - \mu_1 \\ \tilde{E}\{r_2 Y_2/\pi_2\} - \mu_2 \\ \vdots \\ \tilde{E}\{r_T Y_T/\pi_T\} - \mu_T \\ \tilde{E}\{\xi_{T-1}\} \\ \tilde{E}\{\psi_{T-1}\} \end{pmatrix} = \begin{pmatrix} A(\mathcal{X};\theta,\boldsymbol{\phi}) \\ B(\mathcal{X};\boldsymbol{\phi}) \end{pmatrix}.$$

We can obtain $\hat{\theta}$ by minimizing $\tilde{Q}$ with respect to θ, where

$$\tilde{Q} = g(\mathcal{X};\theta,\hat{\boldsymbol{\phi}})'[Var\{g(\mathcal{X};\theta,\boldsymbol{\phi})\}]^{-1} g(\mathcal{X};\theta,\hat{\boldsymbol{\phi}}),$$

which is equivalent to minimizing $A(\mathcal{X};\theta,\hat{\boldsymbol{\phi}})[Var\{A(\mathcal{X};\theta,\hat{\boldsymbol{\phi}})\}]^{-1}A(\mathcal{X};\theta,\hat{\boldsymbol{\phi}})$, similar to our discussion in the $T=1$ case. Notice that the solution to $\tilde{Q}$ would not change, if we rearrange $B(\mathcal{X};\boldsymbol{\phi})$ as

$$B(\mathcal{X};\boldsymbol{\phi}) = \tilde{E}\begin{pmatrix} r_0 u_1 \begin{pmatrix} \frac{L_0}{\pi_0} \\ p_1 L_0 \end{pmatrix} \\ r_1 u_2 \begin{pmatrix} \frac{L_1}{\pi_1} \\ p_2 L_1 \end{pmatrix} \\ \vdots \\ r_{t-1} u_T \begin{pmatrix} \frac{L_{T-1}}{\pi_{T-1}} \\ p_T L_{T-1} \end{pmatrix} \end{pmatrix}.$$

$\tilde{Q}$ can be written as $\tilde{Q} = \tilde{Q}_1 + \tilde{Q}_2$, where $\tilde{Q}_2 = B(\mathcal{X};\hat{\boldsymbol{\phi}})'[Var\{B(\mathcal{X};\boldsymbol{\phi})\}]^{-1}B(\mathcal{X};\hat{\boldsymbol{\phi}})$ and $\tilde{Q}_1$ is

$$\begin{aligned}\tilde{Q}_1 = \{&A(\mathcal{X};\theta,\hat{\boldsymbol{\phi}}) - Cov\{A(\mathcal{X};\theta,\boldsymbol{\phi}), B(\mathcal{X};\boldsymbol{\phi})\}[Var\{B(\mathcal{X};\boldsymbol{\phi})\}]^{-1}B(\mathcal{X};\hat{\boldsymbol{\phi}})\} \\ &\times [Var\{A(\mathcal{X};\theta,\boldsymbol{\phi})\} - Cov\{A(\mathcal{X};\theta,\boldsymbol{\phi}), B(\mathcal{X};\boldsymbol{\phi})\}[Var\{B(\mathcal{X};\boldsymbol{\phi})\}]^{-1} \\ &\times Cov\{B(\mathcal{X};\boldsymbol{\phi}), A(\mathcal{X};\theta,\boldsymbol{\phi})\}]^{-1} \\ &\times \{A(\mathcal{X};\theta,\hat{\boldsymbol{\phi}}) - Cov\{A(\mathcal{X};\theta,\boldsymbol{\phi}), B(\mathcal{X};\boldsymbol{\phi})\}[Var\{B(\mathcal{X};\boldsymbol{\phi})\}]^{-1}B(\mathcal{X};\hat{\boldsymbol{\phi}})\}.\end{aligned}$$

Now consider $Var\{B(\mathcal{X};\boldsymbol{\phi})\}$, which would be a matrix of diagonal blocks, that is,

$$Var\{B(\mathcal{X};\boldsymbol{\phi})\} = diag\left[Var\left\{r_0 u_0 \begin{pmatrix} \frac{L_0}{\pi_0} \\ p_1 L_0 \end{pmatrix}\right\}, \ldots, Var\left\{r_{t-1} u_T \begin{pmatrix} \frac{L_{T-1}}{\pi_{T-1}} \\ p_T L_{T-1} \end{pmatrix}\right\}\right].$$

On the other hand, if we look at $Cov\{A(\mathcal{X};\theta,\boldsymbol{\phi}), B(\mathcal{X};\boldsymbol{\phi})\}$, it is equal to the following lower triangular matrix

$$\begin{pmatrix} Cov\{\tilde{E}(r_0X/\pi_0), B(\mathcal{X};\boldsymbol{\phi})\} \\ Cov\{\tilde{E}(r_1Y_1/\pi_1), B(\mathcal{X};\boldsymbol{\phi})\} \\ Cov\{\tilde{E}(r_2Y_2/\pi_2), B(\mathcal{X};\boldsymbol{\phi})\} \\ \vdots \\ Cov\{\tilde{E}(r_TY_T/\pi_T), B(\mathcal{X};\boldsymbol{\phi})\} \end{pmatrix} = \begin{pmatrix} O \\ Cov\left\{\tilde{E}(r_1Y_1/\pi_1), \tilde{E}\begin{pmatrix}\xi_0\\ \psi_0\end{pmatrix}\right\} \quad O \\ Cov\left\{\tilde{E}(r_2Y_2/\pi_1), \tilde{E}\begin{pmatrix}\xi_1\\ \psi_1\end{pmatrix}\right\} \quad O \\ \vdots \\ Cov\left\{\tilde{E}(r_TY_T/\pi_1), \tilde{E}\begin{pmatrix}\xi_{T-1}\\ \psi_{T-1}\end{pmatrix}\right\} \quad O \end{pmatrix}.$$

Therefore, μ_t can be estimated by solving

$$\tilde{E}\{r_tY_t/\hat{\pi}_t\} - \mu_t - \hat{C}ov\left\{\tilde{E}(r_TY_T/\pi_1), \tilde{E}\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}\end{pmatrix}\right\} Var^{-1}\left\{r_{t-1}u_t\begin{pmatrix}\frac{L_{t-1}}{\pi_{t-1}}\\ p_tL_{t-1}\end{pmatrix}\right\}\tilde{E}\begin{pmatrix}\hat{\xi}_{t-1}\\ \hat{\psi}_{t-1}\end{pmatrix},$$

which is the same as minimizer of Q_t in (5.29).

5.A.4 Sketch of Proof for Corollary 5.1

Proof. With similar arguments to the proof of Theorem 5.1, under conditions (C1)-(C8), we have

$$\begin{aligned} \tilde{E}_A(\hat{\xi}_{t-1}) &= \tilde{E}_A(\xi_{t-1}) - E\{\tilde{E}_A(\xi_{t-1}\psi'_{t-1})|\mathcal{F}_N\}[E\{\tilde{E}_A(\psi_{t-1}\psi'_{t-1}/w)|\mathcal{F}_N\}]^{-1} \\ &\quad \times \tilde{E}_A(\psi_{t-1}/w) + o_p(n^{-1/2}), \\ \tilde{E}_A(r_tY_t/\hat{\pi}_t) &= \tilde{E}_A(r_tY_t/\pi_t) - E[\tilde{E}_A\{(r_tY_t/\pi_t)\psi'_{t-1}\}|\mathcal{F}_N][E\{\tilde{E}_A(\psi_{t-1}\psi'_{t-1}/w)|\mathcal{F}_N\}]^{-1} \\ &\quad \times \tilde{E}_A(\psi_{t-1}/w) + o_p(n^{-1/2}). \end{aligned}$$

Note that $E\{(r_{it}/p_{it}-1)r_{i,t-1}|A,\mathcal{F}_N\} = E[E\{(r_{it}/p_{it}-1)r_{i,t-1}|r_{i,t-1},A,\mathcal{F}_N\}|A,\mathcal{F}_N] = 0$, then $E\{\tilde{E}_A(\xi_{t-1})|A,\mathcal{F}_N\} = 0, E\{\tilde{E}_A(\psi_{t-1}/w)|A,\mathcal{F}_N\} = 0$, and

$$\begin{aligned} Cov\{\tilde{E}_A(\xi_{t-1}), \tilde{E}_A(\psi_{t-1}/w)|\mathcal{F}_N\} &= E\{\tilde{E}_A(\xi_{t-1}\psi_{t-1})'|\mathcal{F}_N\} \\ Var\{\tilde{E}_A(\psi_{t-1})/w|\mathcal{F}_N\} &= E\{\tilde{E}_A(\psi_{t-1}\psi'_{t-1}/w)|\mathcal{F}_N\} \\ Cov\{\tilde{E}_A(r_tY_t/\pi_t), \tilde{E}_A(\psi_{t-1}/w)|\mathcal{F}_N\} &= E[\tilde{E}_A\{(r_tY_t/\pi_t)\psi'_{t-1}\}|\mathcal{F}_N]. \end{aligned}$$

The rest of this proof would follow similarly from the proof of Theorem 5.1. One important step is to calculate $Var[\tilde{E}_A\{(\xi'_{t-1}, \psi'_{t-1}/w)'\}|\mathcal{F}_N]$ and $Cov[\tilde{E}_A(r_tY_t/\pi_t), \tilde{E}_A\{(\xi'_{t-1}, \psi'_{t-1}/w)'\}|\mathcal{F}_N]$.

$$Var\left\{\tilde{E}_A\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}/w\end{pmatrix}\middle|\mathcal{F}_N\right\} = Var\left[E\left\{\tilde{E}_A\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}/w\end{pmatrix}\middle|A,\mathcal{F}_N\right\}\middle|\mathcal{F}_N\right] + E\left[Var\left\{\tilde{E}_A\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}/w\end{pmatrix}\middle|A,\mathcal{F}_N\right\}\middle|\mathcal{F}_N\right].$$

We only have to calculate the second term as the first term is 0. For the second term, notice that

$$\begin{aligned}
&Var\left\{\tilde{E}_A\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}/w\end{pmatrix}\middle|A,\mathcal{F}_N\right\}\\
&= \frac{1}{N^2}\sum_{i\in A} w_i^2 Var\left\{\begin{pmatrix}\xi_{i,t-1}\\ \psi_{i,t-1}/w_i\end{pmatrix}\middle|A,\mathcal{F}_N\right\}\\
&= \begin{pmatrix} N^{-2}\sum_{i\in A}\omega_i^2 Var(\xi_{i,t-1}|A,\mathcal{F}_N) & N^{-2}\sum_{i\in A}\omega_i Cov(\xi_{i,t-1},\psi_{i,t-1}|A,\mathcal{F}_N)\\ N^{-2}\sum_{i\in A}\omega_i Cov(\psi_{i,t-1},\xi_{i,t-1}|A,\mathcal{F}_N) & N^{-2}\sum_{i\in A} Var(\psi_{i,t-1}|A,\mathcal{F}_N)\end{pmatrix}.
\end{aligned}$$

Again $Var(\xi_{i,t-1}|A,\mathcal{F}_N)$ can be written as a matrix of diagonal blocks such that it is equal to

$$diag[Var\{(r_{i1}/p_{i1}-1)r_{i0}L_{i,0}/\pi_{i,0}|A,\mathcal{F}_N\},\dots,Var\{(r_{it}/p_{it}-1)r_{i,t-1}L_{i,t-1}/\pi_{i,t-1}|A,\mathcal{F}_N\}],$$

where

$$Var\{(r_{ij}/p_{ij}-1)r_{i,j-1}L_{i,j-1}/\pi_{i,j-1}|A,\mathcal{F}_N\} = \frac{L_{i,j-1}L'_{i,j-1}}{\pi^2_{i,j-1}}(1/p_{ij}-1)\pi_{i,j-1}.$$

Other related terms can be obtained in a similar fashion. Thus

$$Var\left\{\tilde{E}_A\begin{pmatrix}\xi_{t-1}\\ \psi_{t-1}/w\end{pmatrix}\middle|\mathcal{F}_N\right\} = \begin{pmatrix}\tilde{V}_{LL,t} & \tilde{V}_{LS,t}\\ \tilde{V}_{SL,t} & \tilde{V}_{SS,t}\end{pmatrix},$$

where

$$\begin{aligned}
\tilde{V}_{LL,t} &= N^{-2}\sum_{i=1}^N \omega_i diag\left\{\frac{L_{i,0}L'_{i,0}}{\pi^2_{i,0}}(1/p_{i1}-1)\pi_{i,0},\dots,\frac{L_{i,t-1}L'_{i,t-1}}{\pi^2_{i,t-1}}(1/p_{it}-1)\pi_{i,t-1}\right\},\\
\tilde{V}_{LS,t} &= N^{-2}\sum_{i=1}^N diag\left\{\frac{p_{i1}L_{i,0}L'_{i,0}}{\pi_{i,0}}(1/p_{i1}-1)\pi_{i,0},\dots,\frac{p_{it}L_{i,t-1}L'_{i,t-1}}{\pi_{i,t-1}}(1/p_{it}-1)\pi_{i,t-1}\right\},\\
\tilde{V}_{SS,t} &= N^{-2}\sum_{i=1}^N \omega_i^{-1} diag\left\{p^2_{i1}L_{i,0}L'_{i,0}(1/p_{i1}-1)\pi_{i,0},\dots,p^2_{it}L_{i,t-1}L'_{i,t-1}(1/p_{it}-1)\pi_{i,t-1}\right\}.
\end{aligned}$$

$Cov[\tilde{E}_A(r_tY_t/\pi_t), \tilde{E}_A\{(\xi'_{t-1}, \psi'_{t-1}/w)'\}|\mathcal{F}_N] = (\tilde{V}'_{YL}, \tilde{V}'_{YS})'$, where

$$\tilde{V}_{YL} = N^{-2}\sum_{i=1}^{N}\omega_i Y_{i,t}\left\{(1/p_{i1}-1)L'_{i,0}/\pi_{i0}, \ldots, (1/p_{it}-1)L'_{i,t-1}/\pi_{i,t-1}\right\}'$$
$$\tilde{V}_{YS} = N^{-2}\sum_{i=1}^{N}Y_{i,t}\left\{(1-p_{i1})L'_{i,0}, \ldots, (1-p_{it})L'_{i,t-1}\right\}'.$$

Similarly to the diagonal block-wise technique used in the proof of Theorem 5.1, we obtain the optimal $B_t* = ({B^*_{1t}}', \ldots, {B^*_{tt}}')'$, where

$$\begin{aligned}
B^*_{j,t} &= \{\tilde{V}_{LL,t}(j) - \tilde{V}_{LS,t}(j)\tilde{V}^{-1}_{SS,t}(j)\tilde{V}_{SL,t}(j)\}^{-1}\{\tilde{V}_{LY,t}(j) - \tilde{V}_{LS,t}(j)\tilde{V}_{SS,t}(j)^{-1}\tilde{V}_{SY,t}(j)\} \\
&= (I_{dim(L_{j-1})}, 0)\begin{pmatrix}\tilde{V}_{LL}(j) & \tilde{V}_{LS}(j) \\ \tilde{V}_{LS}(j) & \tilde{V}_{SS}(j)\end{pmatrix}^{-1}\begin{pmatrix}\tilde{V}_{LY}(j) \\ \tilde{V}_{SY}(j)\end{pmatrix} \\
&= (I_{dim(L_{j-1})}, 0)\left[\tilde{E}\left\{w\pi_{j-1}\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1}/w\end{pmatrix}\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1}/w\end{pmatrix}'\right\}\right]^{-1} \\
&\quad\times\tilde{E}\left\{w\left(\frac{1}{p_j}-1\right)\begin{pmatrix}\frac{1}{\pi_{j-1}}L_{j-1} \\ p_jL_{j-1}/w\end{pmatrix}Y_t\right\}.
\end{aligned}$$

The consistency of the estimator in (5.44) naturally follows. □

References

Beaumont, J.-F. and Bocci, C. (2009). Variance estimation when donor imputation is used to fill in missing values. *Canadian Journal of Statistics*, 37:400–416.

Bollinger, C. R. and David, M. H. (1997). Modeling discrete choice with response error: Food stamp participation. *Journal of the American Statistical Association*, 92:827–835.

Bollinger, C. R. and David, M. H. (2001). Estimation with response error and nonresponse: Food-stamp participation in the sipp. *Journal of Business & Economic Statistics*, 19:129–141.

Cao, W., Tsiatis, A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96:723–724.

Fay, R. E. (1992). When are inferences from multiple imputation valid? In *Proc. Survey Res. Meth. Sect.*, pages 227–232. Washington, DC: American Statistical Association.

Flanders, W. D. and Greenland, S. (1991). Analytic methods for two-stage case-control studies and other stratified designs. *Statistics in Medicine*, 10:739–747.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Hawkes, D. and Plewis, I. (2006). Modelling non-response in the national child development study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169:479–491.

Henmi, M. and Eguchi, S. (2004). A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, 91:929–941.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35:501–514.

Kim, J. K., Navarro, A., and Fuller, W. A. (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association*, 101(473):312–320.

Kim, J. K. and Rao, J. N. K. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, 96(4):917–932.

Korinek, A., Mistiaen, J. A., and Ravallion, M. (2007). An econometric method of correcting for unit nonresponse bias in surveys. *Journal of Econometrics*, 136:213–235.

Little, R. J. and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, 31:161–168.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121.

Manski, C. F. and Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica*, 45:1977–1988.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1 edition.

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*, 10(2):475–478.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics*, 10:462–474.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.

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

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82:387–394.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93:1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–590.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94(445):254–265.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101:1619–1637.

Wun, L.-M., Ezzati-Rice, T. M., Diaz-Tena, N., and Greenblatt, J. (2007). On modelling response propensity for dwelling unit (du) level non-response adjustment in the medical expenditure panel survey (meps). *Statistics in Medicine*, 26:1875–1884.