Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 1, 2019 (New Series), pp 53– 61

Theoretical Comparisons of Estimators of Finite Population Proportion Under Simple Random Sampling

Sumanta Adhya[1] **and Tathagata Banerjee**[2]
[1]*West Bengal State University, India.*
[2]*Indian Institute of Management Ahmedabad, India.*

Received: March 26, 2018; Revised: July 09, 2018; Accepted: July 15, 2018

Abstract

We consider the classical survey problem of estimation of finite population proportions based on a polychotomous response variable when data on an auxiliary variable is known for all units in the finite population. Under simple random sampling different model and design-based estimators are compared theoretically and it is shown that model-based estimator performs more efficiently under mild conditions.

Key words: Model-based; Model-assisted; Anticipated variance.

---

1 Introduction

Finite population proportion estimation is a common problem in survey sampling. Suppose variable of interest has $p(\geq 2)$ categories. Let us consider a finite population of N identifiable units defined by $U = \{1,..., N\}$. Then we want to estimate the finite population proportions

$$P_N = (P_1,..., P_p)^T \equiv N^{-1}\sum_U d_i ,$$

of a polychotomous response variable $d = (d_1,..., d_p)^T$ with its value $d_i = (d_{i1},..., d_{ip})^T$ for the i -th population unit, $i \in U$, where $d_{ih} = 1$ if i -th unit belongs to the h-th category and 0 otherwise. Based on a simple random sample (SRS) $S(\subset U)$ of n distinct units the classical design-based estimator is the Horvitz-Thompson (HT) estimator (Horvitz and Thompson (1952)) given by the sample mean

$$\hat{P}_{h,HT} = n^{-1}\sum_S d_{ih} ,\ h = 1,..., p .$$

Often apart from data on the response variable (say, $d_S = \{d_i; i \in S\}$), unit level data on some auxiliary variable x is also available from some other sources. Then $x_U = \{x_i; i \in U\}$ is called complete auxiliary information. A fundamental question in finite population inference is how to make use of the complete auxiliary information effectively at the estimation stage. To this end, two distinct approaches exist, viz., design-based and model-based estimation. Both the estimation procedures employ a regression model (called superpopulation model) relating the response

---

Corresponding Author: Tathagata Banerjee
Email: tathagata.bandyopadhyay@gmail.com

variable and the covariate χ to include covariate information by predicting responses for the non-sample units $\bar{S}(=U-S)$. Under design-based paradigm the basis of inference is randomization due to random sampling from finite population U ; whereas in model-based approach inference is based on sampling distribution from infinite superpopulation model. These two distinct approaches are reconciled by many authors from different aspects (e.g., Smith (1976); Sarndal et al. (1978); and Little (2004) and references therein). The most sharply drawn difference among these estimators from data analytic point of view is in terms of robustness. Grossly speaking, when the working model approximates the true model, model-based estimator's exhibit higher precision; while misspecification of working model may lead to a substantial model bias. In contrast, a class of design-based estimators, called model-assisted estimators, are approximately unbiased with respect under repeated sampling irrespective of the validity of the working models, but show enhanced precision when the working model approximates the true model well.

To compare theoretically these two kinds of estimators, *viz*., the design and model-based estimators representing two different paradigms, Isaki and Fuller (1982) introduce a criterion, called anticipated variance. Anticipated variance is defined as the variance of an estimator with respect to the sampling design and the superpopulation model. In this paper we provide theoretical comparisons of different model-based and model-assisted estimators of finite population proportion P_N by extending the definition of anticipated variance for approximately model and design-unbiased estimators. The motivation of this work is traced back to our earlier work Adhya et al. (2011). Through an extensive simulation studies it has been observed that (*i*) model misspecification has an adverse effect on the performance of model-based estimator, still its performance is comparable with the model-assisted estimators; (*ii*) surprisingly, "the model-assisted estimators are found to be insensitive to the model choice, and hence fail to incorporate auxiliary information well Adhya et al. (2011); p 799. These findings indicate that the design-based estimators of population proportion may not be a good choice. It prompts us to study how well model-based estimators perform when the model is assumed to be true since the possibility of gross model misspecification can be avoided by implementing a simple graphical method to select a working model which approximates the true model well (Adhya et al., 2011; p. 796).

The rest of the paper is outlined as follows. In section 2, we define multinomial logit type model in superpopulation and briefly introduce the model-based and the model-assisted estimators of population proportions. We provide some theoretical comparisons among the estimators in section 3. Finally, we conclude in section 4.

2 Competing Estimators

Here we consider the superpopulation model which assumes observations (d_i, x_i)'s are independent with

$$P(d_{ih}=1|x_i)=\pi_h(x_i;\beta), \qquad (1)$$

$h=1,\ldots,p$, where $\pi_h(x_i;\beta)\equiv\pi_{ih}(\beta)=\exp\{g_h(x_i;\beta_h)\}[1+\exp\{\sum_{u=1}^{p} g_u(x_i;\beta_u)\}]^{-1}$,

$\sum_h \pi_{ih}(\beta)=1$, $\beta_h=(\beta_{h1},\ldots,\beta_{ha_h})^T$, $a_h\ge 1$, $h=1,\ldots,p-1$, and $\beta=(\beta_1^T,\ldots,\beta_{p-1}^T)^T$, and $g_h(.;\beta_h)$, $h=1,\ldots,p-1$, are known but arbitrary parametric functions. Based on the sample data $\{d_S, x_S\}$, let $\hat{\beta}_S$ be the maximum likelihood estimator (MLE) of model parameter β , obtained by maximizing the sample log-likelihood function

$$l_S(\beta \mid d_S, x_S) = \sum_S l_i(\beta) \equiv \sum_S \sum_{h=1}^{p} d_{ih} \ln\{\pi_{ih}(\beta)\} .$$

For details we refer to Adhya et al. (2011). We consider three estimators of P_N for comparison. First, we consider the model-based (MB) estimator (Royall (1970, 1976)) (say, $\hat{P}_{h,M}$) which is the estimated best predictor of finite population proportion P_h given the sample S and the observed data $\{d_S, x_U\}$ under model (1). This is given by

$$\hat{P}_{h,M} = N^{-1}[\sum_S d_{ih} + \sum_{\bar{S}} \hat{\pi}_{ih}], \tag{2}$$

where the predictor $\hat{\pi}_{ih} = \pi_{ih}(\hat{\beta}_S)$ is the predictor of d_{ih} for $i \in \bar{S}$ Adhya et al. (2011).Next we consider two most commonly used model-assisted estimators when relationship between response and auxiliary variable is nonlinear. These are Generalized difference (GD) estimator (Sarndal (1980)) (say, $\hat{P}_{h,GD}$) and Model-calibrated (MC) estimator Wu and Sitter (2001) (say, $P_{h,MC}$). The GD estimator is given by

$$\hat{P}_{h,GD} = \hat{P}_{h,HT} + N^{-1}\sum_U \hat{\pi}_{ih} - n^{-1}\sum_S \hat{\pi}_{ih} \tag{3}$$

where the bias $\hat{P}_{h,HT} - P_h$ of HT estimator is adjusted by the difference $N^{-1}\sum_U \hat{\pi}_{ih} - n^{-1}\sum_S \hat{\pi}_{ih}$, assuming $\hat{\pi}_{ih}$ is a good proxy of d_{ih} for $i \in U$ (that is, $d_h \approx \hat{\pi}_h$). The MC estimator

$$\hat{P}_{h,MC} = \sum_S w_{ih} d_{ih} ,$$

is given by the set of weights $\{w_{ih}; i \in S\}$ that minimizes the aggregated square distance $\sum_S (w_{ih} - n/N)^2$ between w_{ih} and design weight n/N ($i \in S$) subject to calibration constraints

$$N^{-1}\sum_S w_{ih} = 1, \text{ and } \sum_S w_{ih}\hat{\pi}_{ih} = N^{-1}\sum_U \hat{\pi}_{ih} .$$

Notice that MC estimator reduces to

$$\hat{P}_{h,MC} = \hat{P}_{h,HT} + \hat{B}_h\{N^{-1}\sum_U \hat{\pi}_{ih} - n^{-1}\sum_S \hat{\pi}_{ih}\} , \tag{4}$$

where $\hat{B}_h = \{\sum_S (\hat{\pi}_{ih} - \bar{\hat{\pi}}_{h,S})^2\}^{-1} \sum_S (d_{ih} - \bar{d}_{h,S})(\hat{\pi}_{ih} - \bar{\hat{\pi}}_{h,S})$ is the regression coefficient of d_h on $\hat{\pi}_h$ based on $\{(d_{ih}, \hat{\pi}_{ih}), i \in S\}$, $\bar{d}_{h,S} = n^{-1}\sum_S d_{ih}$, $\bar{\hat{\pi}}_h = n^{-1}\sum_S \hat{\pi}_{ih}$ (Wu and Sitter (2001); eqn. (9)). Here also the bias $\hat{P}_h - \hat{P}_{HT}$ of HT estimator is adjusted by scaled difference $\hat{B}_h\{N^{-1}\sum_U \hat{\pi}_{ih} - n^{-1}\sum_S \hat{\pi}_{ih}\}$, assuming $d_h \propto \hat{\pi}_h$.

3 Asymptotic Efficiency of MB Estimator

In this section, we compare the performance of the $\hat{P}_{h,M}$ with the model-assisted estimators $\hat{P}_{h,GD}$ and $\hat{P}_{h,MC}$ by computing their anticipated variances. For large sample study, we assume the asymptotic set up as considered in Isaki and Fuller (1982) (p. 90). In short, we consider a sequence of finite populations $\{U_\nu\}$ (of size $\{N_\nu\}$), and samples $\{S_\nu(\subset U_\nu)\}$ (of size $\{n_\nu\}$) drawn from the corresponding population using a sequence of sampling designs. As the index $\nu \to \infty$ both N_ν and $n_\nu \to \infty$ such that $n_\nu / N_\nu \to \rho \in (0,1)$. Hereafter, we suppress the index ν to simplify the notations. The model-assisted estimators are asymptotically design-unbiased irrespective of whether the working model is correct or not Wu and Sitter (2001). In other words the estimators are asymptotically equal to P_h under SRS from the same population. The only uncertainty is which sample will be observed, and this uncertainty disappears under repeated sampling. For the model-based predictive estimator on the other hand, the uncertainty is about the distribution of non-sample values. It is shown to be asymptotically model-unbiased for the finite population proportion P_h under standard assumptions Adhya et al. (2011). The predictive estimator is then asymptotically equal to P_h. Though the model-based and the model-assisted approaches represent two entirely different paradigms for judging the performance of the estimators, in practice, however, the estimators are used to estimate the same finite population entity. With this consideration in mind, we ask: are the predictive estimators more efficient than the model-assisted estimators in large samples if one uses the large sample anticipated variance as the performance criterion? This is variance anticipated at the time of sample being constructed which incorporates "both the survey designer's conceptualization of a superpopulation (prior knowledge) and the design". The theoretical results under a simple random sampling using intuitive arguments proved below show that the answer is in affirmative.

From the proof of Theorem 1 of Adhya et al. (2011), notice that for a generic model-assisted estimator $\hat{P}_{h,ma}$ anticipated variance of the prediction error $AV\{\hat{P}_{h,ma} - P_h\}$ is given by

$$AV\{\hat{P}_{h,ma} - P_h\} = E_d V_m\{\hat{P}_{h,ma} - P_h \mid S\} + o(n^{-1}),$$

where $E_d(.)$ and $V_d(.)$ ($E_m(.)$ and $V_m(.)$) denote design (model) expectation and variance respectively. On the other hand for model-based estimator

$$AV\{\hat{P}_{h,N} - P_h\} = E_d V_m\{\hat{P}_{h,N} - P_h \mid S\} + o(n^{-1}).$$

Thus to compare $AV\{\hat{P}_{h,m} - P_h\}$ and $AV\{\hat{P}_{h,ma} - P_h\}$ in large sample we ignore $o(n^{-1})$ terms and compare only large sample anticipated variances $E_d V_m\{\hat{P}_{h,m} - P_h \mid S\}$ and $E_d V_m\{\hat{P}_{h,ma} - P_h \mid S\}$ respectively. For this it is enough to compare conditional model variances $V_m\{. \mid S\}$ uniformly for all samples *S*. Now we state our results concerning $V_m\{. \mid S\}$ in large samples. The proofs are given in the Appendix.

***Theorem* 1.** Under the regularity conditions stated in section 2 of Adhya et al. (2011),

$(i) V_m\{\hat{P}_{h,GD} - P_h \mid S\} \big/ V_m\{\hat{P}_{h,M} - P_h \mid S\} \approx 1 + V_m\{n^{-1}\sum_S (\hat{\pi}_{ih} - d_{ih})\} / V_m\{(N-n)^{-1}\sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih})\}$
in large samples, and
$(ii) V_m\{\hat{P}_{h,GD} - P_h \mid S\} = V_m\{\hat{P}_{h,MC} - P_h \mid S\} + o(n^{-1})$.

From the equality in (*ii*) of Theorem 1, model variances of $\hat{P}_{h,MC}$ and $\hat{P}_{h,GD}$ estimators are asymptotically equal; whereas (*i*) and (*ii*) together imply that $\hat{P}_{h,M}$ is more efficient than $\hat{P}_{h,GD}$ and $\hat{P}_{h,MC}$. Also from (*i*) it is expected that efficiency of that of $\hat{P}_{h,M}$ will be high when N is large compared to n. This follows from the fact that given s, the variance of mean of the differences $(\hat{\pi}_h - d_h)$ based on non-sample units is small compared to variance of mean of the differences based on sample units.

Before stating following corollaries let us define

$\bar{\pi}_h(\beta) = \lim_{n\to\infty} n^{-1}\sum_S \pi_{ih}(\beta)$, $\bar{\pi}'_h(\beta) = \lim_{n\to\infty} n^{-1}\sum_S \partial\pi_{ih}(\beta)/\partial\beta$,
$I_{i_{hh'}}(\beta) = E_m\{-\partial^2 l_i(\beta)/\partial\beta_h\partial\beta_{h'}^T\}$, $\bar{I}(\beta) = \lim_{n\to\infty} n^{-1}\sum_S (I_{ihh'}(\beta))$, and
$\bar{v}_{hh}(\beta) = \lim_{n\to\infty} (N-n)^{-1}\sum_{\bar{S}} \pi_{ih}(\beta)\{1-\pi_{ih}(\beta)\}$,
where details of the expressions are given in Adhya et al. (2011), and the limits are defined with respect to asymptotic density of χ Chambers et al. (1992).

***Corollary* 1**. If the limiting sampling fraction ρ is less than equal to 0.5 then the asymptotic efficiency of $\hat{P}_{h,M}$ relative to $\hat{P}_{h,GD}$ is at the least equal to 2.

Proof. From Theorem 1, we obtain

$$V_m\{\hat{P}_{h,GD} - P_h \mid S\} \big/ V_m\{\hat{P}_{h,M} - P_h \mid S\} \geq 1 + c_S / c_{\bar{S}}\ , \tag{5}$$

where $c_S = V_m\{n^{-1}\sum_S (\hat{\pi}_{ih} - d_{ih})\}$ and $c_{\bar{S}} = V_m\{(N-n)^{-1}\sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih})\}$. After some algebra, neglecting the terms of order $o(n^{-1})$, we obtain $c_S \approx 3n^{-1}\bar{\pi}'_h(\beta)^T \bar{I}^{-1}(\beta)\bar{\pi}'_h(\beta) + n^{-1}\bar{v}_{hh}(\beta)$ and $c_{\bar{S}} \approx n^{-1}\bar{\pi}'_h(\beta)^T \bar{I}^{-1}(\beta)\bar{\pi}'_h(\beta) + n^{-1}\{\rho/(1-\rho)\}\bar{v}_{hh}(\beta)$, thus yielding

$$c_s = c_{\bar{S}} + 2n^{-1}\bar{\pi}'_h(\beta)^T \bar{I}^{-1}(\beta)\bar{\pi}'_h(\beta) + n^{-1}\{1 - \frac{\rho}{1-\rho}\}\bar{v}_{hh}(\beta).$$

So (5) entails

$$\frac{V_m\{\hat{P}_{h,GD} - P_h \mid S\}}{V_m\{\hat{P}_{h,M} - P_h \mid S\}} \approx 2 + \frac{2n^{-1}\bar{\pi}'_h(\beta)^T \bar{I}^{-1}(\beta)\bar{\pi}'_h(\beta) + n^{-1}\{(1-2\rho)/(1-\rho)\}\bar{v}_{hh}(\beta)}{c_{\ddot{S}}}$$

and hence the result.

***Corollary* 2.** As $\rho \downarrow 0$, the asymptotic efficiency of $\hat{P}_{h,m}$ relative to $\hat{P}_{h,GD}$ increases to ∞.

This follows immediately from (5) by noting that $c_S = O(n^{-1})$ and $c_{\bar{S}} = O((N-n)^{-1})$.

4 Concluding Remarks

Adhya et al. (2011) show that for polychotomous responses the model-assisted estimators viz. GD estimator and MC estimator of finite population proportion are model insensitive, and hence they fail to incorporate unit level auxiliary information. Here we establish theoretically that Adhya et al. (2011)'s MB estimator is asymptotically more efficient than the model-assisted estimators under simple random sampling when sampling fraction is relatively small and the assumed model is true. Since for multinomial-type logit model a simple graphical procedure as described in Adhya et al. (2011) leads to the selection of a good working model, in the light of this fact we prefer the model-based estimator $\hat{P}_{h,M}$ to model-assisted estimators $\hat{P}_{h,GD}$ and $\hat{P}_{h,MC}$ under simple random sampling if the sampling fraction is small. In future, we will consider similar study for other sampling designs.

References

Adhya, S., Banerjee, T., and Chattopadhyay, G. (2011). Inferences on polychotomous responses infinite populations. *Scandinavian Journal of Statistics*, **38**, 788-800.

Chambers, R.L., Dorfman, A.H., and Hall, P. (1992). Properties of estimators of finite population distribution function. *Biometrika*, **79**, 577-582.

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.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, **77**, 89-96.

Little, R.J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, **99**, 546-556.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression model. *Biometrika*, **57**, 377-387.

Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, **71**, 657-664.

Sarndal, C.E. (1978). Design-based and model-based inference in survey sampling (with discussions). *Scandinavian Journal of Statistics*, **5**, 27-52.

Sarndal, C.E. (1980). On π -inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, **67**, 639-650.

Smith T.M.F. (1976). The foundations of survey sampling: A review (with discussions). *Journal of the Royal Statistical Society*, **A139**, 183-204.

Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, **96**, 185-193.

Appendix: Proof of Theorem 1

Proof of part (i). Rewrite the estimators $\hat{P}_{h,M}$, $\hat{P}_{h,GD}$ and $\hat{P}_{h,MC}$ ((2)-(4)) as:

$$\hat{P}_{h,m} = P_h + w_{\bar{S}} \sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih}) ,$$
$$\hat{P}_{h,GD} = P_h + w_S \sum_S (\hat{\pi}_{ih} - d_{ih}) + w_{\bar{S}} \sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih}) , \text{ and} \quad \text{(A1)}$$
$$\hat{P}_{h,MC} = P_h + \sum_U w_i (\hat{B}_h \hat{\pi}_{ih} - d_{ih}) ,$$

where $w_i = w_S \ for\, i \in S$ and $w_{\bar{S}} \ for\, i \in \bar{S}$. For simplicity of presentation we drop the notation conditioning on S from model expectations and model variances.

Now, (A1) entails

$$V_m(\hat{P}_{h,GD} - P_h) = V_m(\hat{P}_{h,m} - P_h) + V_m\{w_S \sum_{i \in S} (\hat{\pi}_{ih} - d_{ih})\}$$
$$+ 2 w_S w_{\bar{S}} Cov_m\{\sum_S (\hat{\pi}_{ih} - d_{ih}), \sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih})\}, \quad \text{(A2)}$$

where $Cov_m(.)$ is the model-based covariance. Letting $\pi_{ih}(\beta) = \pi_{ih}$ and $\pi'_{ih}(\beta) = \pi'_{ih}$, we have

$$2 w_S w_{\bar{S}} Cov_m\{\sum_S (\hat{\pi}_{ih} - d_{ih}), \sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih})\}$$
$$= w_S w_{\bar{S}} Cov_m\{\sum_S (\pi_{ih} - d_{ih}) + (\hat{\beta}_S - \beta)^T \sum_S \pi'_{ih}, \sum_{\bar{S}} (\pi_{ih} - d_{ih}) + (\hat{\beta}_S - \beta)^T \sum_{\bar{S}} \pi'_{ih}\} + o(n^{-1}) . \text{(A3)}$$

Letting $u_{ih} = g'_{ih}(\beta_h) = \partial g(x_i; \beta_h) / \partial \beta_h$, $S_h(\beta) = \sum_S u_{ih}(d_{ih} - \pi_{ih})$, $S(\beta) = (S_1(\beta)^T, ..., S_{p-1}(\beta)^T)^T$ is the score function based on sample likelihood $l_S(\beta \mid d_S, x_S)$ and from section 2 of Adhya et al. (2011),

$$\hat{\beta}_S - \beta = \bar{I}^{-1}(\beta)\{n^{-1} S(\beta)\} + o_P(n^{-1}), \text{ and } E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\} = n^{-1} \bar{I}^{-1}(\beta) + o(n^{-1}) .$$

Thus (A3) reduces to

$$2 w_S w_{\bar{S}} [(\sum_S \pi'_{ih})^T \{n^{-1} \bar{I}^{-1}(\beta) + o(n^{-1})\} (\sum_S \pi'_{ih})$$
$$- Cov_m\{\sum_S (d_{ih} - \pi_{ih}), n^{-1}(\sum_S u_{i1}^T (d_{i1} - \pi_{i1}), ..., \sum_S u_{ip-1}^T (d_{ip-1} - \pi_{ip-1})) \bar{I}^{-1}(\beta) (\sum_{\bar{S}} \pi'_{ih})\}] + o(n^{-1}).$$

After some algebra it again reduces to

$$2n^{-1} w_S w_{\bar{S}} [(\sum_S \pi'_{ih})^T \bar{I}^{-1}(\beta)(\sum_S \pi'_{ih}) - (\sum_S \pi'_{ih}) \bar{I}^{-1}(\beta)(\sum_{\bar{S}} \pi'_{ih})] + o(n^{-1}) = o(n^{-1}) . \quad \text{(A4)}$$

For large n, (A2) and (A4) entail

$$V_m(\hat{P}_{h,GD} - P_h) \approx V_m(\hat{P}_{h,M} - P_h) + V_m\{w_S \sum_S (\hat{\pi}_{ih} - d_{ih})\}$$
$$= (1 - f)^2 [V_m\{(N - n)^{-1} \sum_{\bar{S}} (\hat{\pi}_{ih} - d_{ih})\} + V_m\{n^{-1} \sum_S (\hat{\pi}_{ih} - d_{ih})\}] .$$

Thus we obtain

$$V_m(\hat{P}_{h,GD} - P_h) \big/ V_m(\hat{P}_{h,M} - P_h) \approx 1 + V_m\{n^{-1} \sum_{i \in S} (\hat{\pi}_{ih} - d_{ih})\} / V_m\{(N - n)^{-1} \sum_{i \in \bar{S}} (\hat{\pi}_{ih} - d_{ih})\} .$$

Proof of part (*ii*). We note that from (4) after linearization $\hat{B}_h \equiv \hat{B}_h(\hat{\beta}_S)$ can be written as

$$\hat{B}_h = B_{0h} + \frac{B_{1h}}{\sqrt{n}} + \frac{B_{2h}}{n} + o_P(n^{-1}), \tag{A5}$$

where B_{0h}, B_{1h} and B_{2h} are stochastically bounded. Thus we have

$$\hat{P}_{h,MC} = P_h + \sum_U w_i(B_{0h}\hat{\pi}_{ih} + n^{-1/2}B_{1h}\hat{\pi}_{ih} + n^{-1}B_{2h}\hat{\pi}_{ih} - d_{ih} + o_P(n^{-1}))$$

$$= P_h + B_{0h}\sum_U w_i\hat{\pi}_{ih} - \sum_U w_i d_{ih} + \frac{B_{1h}}{\sqrt{n}}\sum_U w_i\hat{\pi}_{ih} + O_P(n^{-1}). \tag{A6}$$

Now, we find the expressions for B_{0h} and B_{1h} in (A6). Letting $\pi''_{ih}(\beta) = \pi''_{ih}$, note that

$$\hat{\pi}_{ih} - \pi_{ih} = (\hat{\beta}_S - \beta)^T \pi'_{ih} + (1/2)(\hat{\beta}_S - \beta)^T \pi''_{ih}(\hat{\beta}_S - \beta) + o_P(n^{-1}).$$

Letting $\bar{\pi}_{h,S} = n^{-1}\sum_S \pi_{ih}$, $\bar{\pi}'_{h,S} = n^{-1}\sum_S \pi'_{ih}$, and $\bar{\pi}''_{h,S} = n^{-1}\sum_S \pi''_{ih}$, above expansion gives

$$n^{-1}\sum_S (d_{ih} - \bar{d}_{h,S})(\hat{\pi}_{ih} - \bar{\hat{\pi}}_{h,S})$$

$$= n^{-1}\sum_S (d_{ih} - \bar{d}_{h,S})(\pi_{ih} - \bar{\pi}_{h,S}) + (\hat{\beta} - \beta)^T n^{-1}\sum_S (d_{ih} - \bar{d}_{h,S})(\pi'_{ih} - \bar{\pi}'_{h,S})$$

$$+ (1/2)(\hat{\beta}_S - \beta)^T n^{-1}\sum_S (\pi''_{ih} - \bar{\pi}''_{h,S})(\hat{\beta}_S - \beta) + o_P(n^{-1}), \tag{A7}$$

where

$$n^{-1}\sum_S (\hat{\pi}_{ih} - \bar{\hat{\pi}}_{h,S})^2 = n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})^2 + (\hat{\beta}_S - \beta)^T n^{-1}\sum_S (\pi'_{ih} - \bar{\pi}'_{h,S})(\pi'_{ih} - \bar{\pi}'_{h,S})^T(\hat{\beta}_S - \beta)$$

$$+ 2(\hat{\beta}_S - \beta)^T n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})(\pi'_{ih} - \bar{\pi}'_{h,S})^T$$

$$+ (\hat{\beta} - \beta)^T n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})(\pi''_{ih} - \bar{\pi}''_{h,S})(\hat{\beta}_S - \beta) + o_P(n^{-1})$$

$$= b_{0h} + \frac{b_{1h}}{\sqrt{n}} + \frac{b_{2h}}{n} + o_P(n^{-1}), \tag{A8}$$

where $b_{kh} = O_P(1), k = 0,1,2$. We define $a_{0h} = n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})^2, a_{1h} = (\mathbf{A}8) - a_{0h}$. Note that a_{0h} $= O_P(1)$, $a_{1h} = O_P(n^{-1/2})$. Now, (A7) and (A8) imply

$$\hat{B}_h = (\mathbf{A}7) \times \{n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})^2\}^{-1}(1 + a_{0h}^{-1}a_{1h})^{-1}$$

$$= \{b_{0h} + \frac{b_{1h}}{\sqrt{n}} + \frac{b_{2h}}{n} + o_P(n^{-1})\} \times a_{0h}^{-1}\{1 - a_{0h}^{-1}a_{1h} + (a_{0h}^{-1}a_{1h})^2 + o_P(n^{-1})\}, \text{ say}$$

$$\equiv B_{0h} + \frac{B_{1h}}{\sqrt{n}} + \frac{B_{2h}}{n} + o_P(n^{-1}),$$

where $B_{0h} = \frac{b_{0h}}{a_{0h}} = \frac{n^{-1}\sum_S (d_{ih} - \bar{d}_{h,S})(\pi_{ih} - \bar{\pi}_{h,S})}{n^{-1}\sum_S (\pi_{ih} - \bar{\pi}_{h,S})^2}$. This entails $B_{0h} = 1 + O_P(n^{-1/2})$.

Now, $E_m\{\hat{P}_{h,MC} - P_h\}^2 = E_m\{(B_{0h}\sum_U w_i\hat{\pi}_{ih} - \sum_U w_i d_{ih}) + \frac{B_{1h}}{\sqrt{n}}\sum_U w_i\hat{\pi}_{ih} + O_P(n^{-1})\}^2$, and

$$B_{0h}\sum_U w_i\hat{\pi}_{ih} - \sum_U w_i d_{ih} = \sum_U w_i(\hat{\pi}_{ih} - d_{ih}) + (B_{0h}-1)\sum_U w_i\pi_{ih} + O_P(n^{-1})$$

$$= t_N + (B_{0h}-1)\sum_U w_i\pi_{ih} + O_P(n^{-1}) \text{, say.}$$

Further note that

$$t_N = \sum_U w_i(\hat{\pi}_{ih} - d_{ih}) = \hat{P}_{h,GD} - P_h \,,\, E_m\{t_N \sum_U w_i\hat{\pi}_{ih}\} = E_m\{t_N\}\sum_U w_i\pi_{ih} + O(n^{-1}),$$

where we have $E_m\{t_N\} = O(n^{-1})$. Thus we obtain

$$E_m\{\hat{P}_{h,MC} - P_h\}^2 = E_m\{t_N + (B_{0h}-1)\sum_U w_i\pi_{ih} + \frac{B_{1h}}{\sqrt{n}}\sum_U w_i\hat{\pi}_{ih} + O_P(n^{-1})\}^2$$

$$= E_m\{t_N\}^2 + 2E_m\{(B_{0h}-1)(t_N\sum_U w_{ih}\pi_{ih})\} + B_{1h}\frac{2}{\sqrt{n}}E_m\{t_N\sum_U w_i\hat{\pi}_{ih}\} + o(n^{-1})$$

$$= E_m\{t_N\}^2 + 2E_m\{(B_{0h}-1)(t_N\sum_U w_{ih}\pi_{ih})\} + o(n^{-1})$$

$$= E_m\{t_N\}^2 + 2(1-f)E_m\{(B_{0h}-1)t_N\}\{(N-n)^{-1}\sum_{\bar{S}}\pi_{ih} - n^{-1}\sum_S \pi_{ih})\} + o(n^{-1})$$

$$= E_m\{t_N\}^2 + o(n^{-1})$$

$$= E_m\{\hat{P}_{h,GD} - P_h\}^2 + o(n^{-1}).$$