Pak. J. Statist.
2011 Vol. 27(4), 407-426

ASYMPTOTIC CONSISTENCY UNDER LARGE ENTROPY SAMPLING DESIGNS WITH UNEQUAL PROBABILITIES

Yves G. Berger
University of Southampton, S3RI, SO17 1BJ, Southampton, UK
Email: Y.G.Berger@soton.ac.uk

ABSTRACT

A large part of survey sampling literature is devoted to unequal probabilities sampling designs without replacement. Brewer and Hanif (1983) provided a summary of these sampling designs. The maximum entropy designs is one of them. Consistency results have been proven for the maximum entropy sampling (Hájek, 1964). The aim is to give sufficient conditions under which Hájek (1964) consistency results still hold for large entropy sampling designs which are different from the maximum entropy design. These conditions involve modes of convergence of sampling designs towards the maximum entropy design. We show that these conditions are satisfied for the popular Rao-Sampford (Rao, 1965, Sampford, 1967) design. Our consistency results are applied to the Hájek (1964) simple variance estimator. This estimator does not require joint-inclusion probabilities and can be easily estimated using weighted least squares regression (Berger, 2004, 2005b). Deville (1999) conjectured that this estimator is suitable for any sampling designs (see also Brewer and Donadio, 2003). Our consistency result gives regularity conditions under which this estimator is consistent which justifies Deville's (1999) conjecture.

KEY WORDS:

Consistency; Design-Based Inference; Inclusion Probabilities, π-Estimator, Sample Survey.

Primary AMS: 62D05
Secondary AMS: 94A17

1. INTRODUCTION

Assume that a sample is randomly selected without replacement from a population U. Hence a sample is a random variable which will be denoted by $\mathscr{S}$. An observed sample will be denoted by s. The probability distribution $p(s)$ denotes the probability of selecting the sample s. The distribution $p(s)$ is commonly called the *sampling design* or *design*. In this paper, all the sampling designs considered are uni-stage sampling designs with fixed sample size n.

Let the quantity π_i denote the first-order inclusion probability of unit i; that is, the probabilities to sample unit i. This probability is defined by

$$\pi_i = \sum_s p(s)\delta\{i \in s\} ; \tag{1}$$

where $\delta\{A\} = 1$ if A is true and $\delta\{A\} = 0$ otherwise. Usually the π_i are given quantities which for example, can be proportional to a size variable (Brewer and Hanif, 1983 Chap. 2). In this paper, all the sampling designs considered have the same first-order inclusion probabilities; that is, (1) holds for all the sampling designs considered.

Suppose we wish to estimate a population parameter $\theta(U)$. For example, a total

$$\theta(U) = \sum_{i \in U} y_i , \tag{2}$$

where y_i is the value of a study variable of a unit labelled i. Let $\hat{\theta}(\mathscr{S})$ denote an estimator of $\theta(U)$. An estimate is denoted by $\hat{\theta}(s)$. For example, the π-estimator (Narain, 1951; Horvitz and Thompson, 1952) of a total (2) is defined by

$$\hat{\theta}(\mathscr{S}) = \sum_{i \in \mathscr{S}} \frac{y_i}{\pi_i} .$$

Another parameter of interest that will be considered in Section 7 and 8 is the variance of the π-estimator. In this case, $\theta(U)$ is the variance of the π-estimator and $\hat{\theta}(\mathscr{S})$ is the variance estimator. In Section 7, we will apply our consistency result to the consistency of a variance estimator of the π-estimator.

The concept of entropy has been introduced in Survey sampling theory by Hájek (1981). This concept has been used by Berger (1998a, 1998b, 2005a), Aires (2000), Brewer (2002 p 260), Brewer and Donadio (2003) and Grafström (2010). The *entropy* of a sampling design $p(s)$ is defined by

$$H(p) = -\sum_s p(s) \log(p(s)) ;$$

where we consider that $0 \log(0) = 0$.

In Section 2, we define the asymptotic framework used. In Section 3, we define the maximum entropy design. In Section 4, we define different modes of convergence towards the maximum entropy design and their properties. In Section 5, we show how these modes of convergence can be used to proof consistency results. In Section 6, we propose a set of regularity conditions under which convergences defined in Section 4 hold for the Rao-Sampford (Rao, 1965, Sampford, 1967) design. In Section 7, we apply our results to the consistency of the Hájek (1964) variance estimator. In Section 8, we generalise our results to stratification. A brief simulation study in Section 9 supports our findings.

2. THE ASYMPTOTIC FRAMEWORK

Let $\{\mathscr{S}_t\}$ be a sequence of random variables which denotes random samples selected from a sequence of nested finite populations U_t of size N_t by a sequence of uni-stage probability sampling designs $\{ p_t(s) \}$; where $t = 1, 2, \cdots, \infty$. Let s_t denote an observed sample composed of a fixed number n_t of distinct elements selected from U_t ($n_t < N_t$). Let $\pi_{t,i}$ ($i \in U_t$) be the inclusion probabilities of the sampling design $p_t(s)$. We assume that the inclusion probabilities are such that $\lim_{t\to\infty} d_t = \infty$ where

$$d_t = \sum_{i \in U_t} \pi_{t,i} (1 - \pi_{t,i}) .$$

The assumption $d_t \to \infty$ was first suggested by Hájek (1964). This assumption implies that $n_t \to \infty$ and $N_t \to \infty$, as $d_t < n_t = \sum_{i \in U_t} \pi_{t,i} < N_t$. Note that as we do not assume that $n_t / N_t \to 0$, this asymptotic framework can be used for large sampling fractions.

3. THE MAXIMUM ENTROPY OR REJECTIVE SAMPLING DESIGN

Let $\{ r_t(s) \}$ be the sequence of maximum entropy uni-stage sampling designs (Hájek, 1981, p. 132) defined by

$$r_t(s) = \phi_t \ \delta\{\# s = n_t\} \prod_{i \in s} \frac{p_{t,i}}{1 - p_{t,i}}, \tag{3}$$

where ϕ_t is such that $\sum_s r_t(s) = 1$; $\delta\{\# s = n_t\} = 1$ if s contains n_t units and $\delta\{\# s = n_t\} = 0$ otherwise; and the $p_{t,i}$ as such that the the inclusion probabilities of $r_t(s)$ are given by $\pi_{t,i}$. In order to have a unique set of $p_{t,i}$, we consider that $\sum_{i \in U_t} p_{t,i} = n_t$ (Hájek, 1981, p. 132). Hájek (1964) proposed an approximation for the $p_{t,i}$ (see also Brewer and Hanif, 1983, p. 40). Chen et al. (1994) proposed a method for computing $p_{t,i}$ exactly (see also Aires, 1999, 2000; Traat *et al*., 2004). A review of the different methods to compute the $p_{t,i}$ can be found in Berger and Tillé (2011 p 49).

Hájek (1981 p. 132) showed that

$$H(r_t) \geq H(p_t) \tag{4}$$

for all $p_t(s)$ such that $\sum_s p_t(s) \delta\{i \in s\} = \sum_s r_t(s) \delta\{i \in s\} = \pi_{t,i}$. Hence $r_t(s)$ maximises the entropy.

4. MODE OF CONVERGENCE OF SAMPLING DESIGNS

As the maximum entropy design is rarely implemented in practice, we consider that the actual uni-stage sampling design $p(s)$ implemented is different from the maximum entropy sampling design. We consider that $p(s)$ and the maximum entropy design have the same first-order inclusion probabilities. Let $\{ p_t(s) \}$ be the associated sequence of sampling designs. Consistency (see Section 5) will be based upon the fact that $p_t(s)$

converges to $r_t(s)$. For example, we could used the following definition to define this convergence.

Definition 1:

$p_t(s)$ converge to $r_t(s)$ with respect to their entropy if

$$\lim_{t\to\infty} D(p_t, r_t) = 0\,; \tag{5}$$

where $D(p_t, r_t) = H(r_t) - H(p_t)$.

Note that (4) implies that $D(p_t, r_t) > 0$. Note that $D(p_t, r_t) = 0$ when $p_t(s)$ is the stratified simple random sampling design and $\pi_{t,i} = n_t / N_t$, as $p_t(s) = r_t(s)$ in this situation. Berger (2005a) gave sufficient conditions under which (5) holds under Chao (1982) sampling. These sufficient conditions hold when the inclusion probabilities $\pi_{t,i}$ ($i \le n_t$) tend to zero.

Convergence (5) implies that the entropy of $p(s)$ is large, as $r(s)$ is the maximum entropy design. Thus, this definition excludes low entropy designs such as the systematic sampling designs (Madow and Madow, 1944). However, the randomised systematic sampling design (Madow, 1949) is a high entropy design (Berger, 1998b). Although, there is no formal proof that (5) holds for the randomised systematic sampling design. High Entropy implies that $\pi_{ij} \approx \pi_i \pi_j$ (Hájek, 1981, p. 32), and this ensures efficient variance estimation (Hanurav, 1966; Isaki and Fuller, 1982).

The convergence of $p_t(s)$ towards $r_t(s)$ can also be defined by the following other modes of convergence.

Definition 2:

$p_t(s)$ converge to $r_t(s)$ with respect to the total variation norm if

$$\lim_{t\to\infty} \| p_t - r_t \|_1 = 0\,; \tag{6}$$

where $\| p_t - r_t \|_1 = \sum_s | p_t(s) - r_t(s) |$.

Definition 3:

$p_t(s)$ converge to $r_t(s)$ with respect to the Chi-square distance if

$$\lim_{t\to\infty} \| p_t - r_t \|_2 = 0 \tag{7}$$

where $\| p_t - r_t \|_2 = \sum_s r_t(s)\{1 - p_t(s) / r_t(s)\}^2$.

Definition 4:

$p_t(\mathscr{S}) / r_t(\mathscr{S}) \xrightarrow{r} 1$ *in probability with respect to* $r_t(s)$ *if*

$$\lim_{t\to\infty}\sum_{s} r_t(s)\delta\left\{\left|\frac{p_t(s)}{r_t(s)}-1\right|>\varepsilon\right\}=\lim_{t\to\infty} r_t\left\{\left|\frac{p_t(\mathscr{S})}{r_t(\mathscr{S})}-1\right|>\varepsilon\right\}=0 \qquad (8)$$

for every $\varepsilon > 0$ *; where* $r_t\{\mathscr{B}\}$ *denotes the probability with respect* $r_t(s)$ *to observe and event* $\mathscr{B}$.

The following properties give the relations between the modes of convergence (5) – (8).

Property 1:
Convergence (6) *implies convergence* (8).

Property 2:
Convergence (5) *implies convergence* (6).

Property 3:
Convergence (7) *implies convergence* (6).

The proofs of Properties 1 and 2 can be found in Appendix A and B. Property 3 can be proven by using Cauchy's inequality.

Properties 1,2 and 3 mean that convergence (8) is the weakest convergence. However, we will see in Section 5 that convergence (6) is the key convergence for consistency.

In order to find more relations between the modes of convergence (5) – (8), we need to consider that $p_t(\mathscr{S})/r_t(\mathscr{S})$ is uniformly bounded.

Definition 5:
$p_t(\mathscr{S})/r_t(\mathscr{S})$ *is uniformly bounded if there exist* κ *such that*

$$r_t\left\{\left|\frac{p_t(\mathscr{S})}{r_t(\mathscr{S})}-1\right|<\kappa\right\}=1 \quad \forall t .$$

When $p_t(\mathscr{S})/r_t(\mathscr{S})$ is uniformly bounded, we have the following property.

Property 4:
If $p_t(\mathscr{S})/r_t(\mathscr{S})$ *is uniformly bounded, any mode of convergence implies any other modes of convergence*

The proof is given in Appendix C.

5. ASYMPTOTIC CONSISTENCY

In this section, we define the asymptotic consistency and propose regularity conditions under which an estimator is consistent under a sampling design $p(s)$.

Definition 6:

$\hat{\theta}_t(\mathscr{S})$ is an asymptotically consistent estimator of a population quantity $\theta_t(U_t)$ under the sampling design $p(s)$ if $\forall \varepsilon > 0$

$$\lim_{t\to\infty} p_t\{|\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)| > \varepsilon\} = 0 ;$$

where $p_t\{\mathscr{B}\}$ denote the probability with respect $p_t(s)$ to observed an event $\mathscr{B}$.

Theorem 1 and Corollary 1 give conditions under which $\hat{\theta}_t(\mathscr{S})$ is consistent under a sampling design $p(s)$.

Theorem 1:

If $\hat{\theta}_t(\mathscr{S})$ is asymptotically consistent under the maximum entropy sampling design $r(s)$, then $\hat{\theta}_t(\mathscr{S})$ is asymptotically consistent under $p(s)$ if convergence (6) *holds.*

The proof of Theorem 1 is given in Appendix D. The following corollary follows from Properties 2 and 3.

Corollary 1:

$\hat{\theta}_t(\mathscr{S})$ is consistent under $p(s)$ if convergence (5) *or* (7) *holds and if $\hat{\theta}_t(\mathscr{S})$ is consistent under $r(s)$.*

The following corollary follows from Property 4.

Corollary 2:

If $p_t(\mathscr{S}) / r_t(\mathscr{S})$ is uniformly bounded and if (8) *holds, we have that $\hat{\theta}_t(\mathscr{S})$ is consistent under $p(s)$ if $\hat{\theta}_t(\mathscr{S})$ is consistent under $r(s)$.*

In Section 7, we show that $p_t(\mathscr{S}) / r_t(\mathscr{S})$ is uniformly bounded and (8) holds for the Rao-Sampford design (Rao, 1965; Sampford, 1967).

6. CONVERGENCE OF THE RAO-SAMPFORD DESIGN

The Rao-Sampford sampling design is a popular design used for unequal probability sampling without replacement. The main advantage of the Rao-Sampford design is its simplicity. In the Appendix E and F, we show that $p_t(\mathscr{S}) / r_t(\mathscr{S})$ is uniformly bounded and that (8) holds under condition 1 and 2 below. Hence from corollary 2, $\hat{\theta}_t(\mathscr{S})$ is consistent under Rao-Sampford sampling if $\hat{\theta}_t(\mathscr{S})$ is consistent under maximum entropy sampling.

Condition 1:

$n_t / \tilde{d}_t = O(1)$ where $\tilde{d}_t = \sum_{i\in U_t} p_{t,i}(1 - p_{t,i})$.

Condition 2:

There exist constant κ_m *and* κ_M *such that* $\kappa_m < \hat{d}_t d_t^{-1} < \kappa_M$.

In Appendix G, we show that if $\pi_{t,i} \leq \pi \neq 1$ for all $i \in U_t$, Condition 1 holds. The condition $\pi_{t,i} \leq \pi$ ensures that the $\pi_{t,i}$ are uniformly bounded and that none of the $\pi_{t,i}$ tends to one. This condition may hold in practice, although this condition is stronger that Condition 1.

7. CONSISTENCY OF THE HÁJEK VARIANCE ESTIMATOR

Hájek (1964, p. 1520) proposed a simple estimator for the variance of the π-estimator

$$\hat{Y} = \sum_{i \in \mathscr{S}} \frac{y_i}{\pi_i} .$$

of a total $Y = \sum_U y_i$. Berger (2004, 2005b) showed that the Hájek (1964, p. 1520) variance estimator equals

$$\hat{\text{var}}(\hat{Y})_H = \sum_{i \in \mathscr{S}} c_i \, \hat{e}_i^2 \tag{9}$$

where $c_i = n(n-1)^{-1}(1-\pi_i)$. The $\hat{e}_i$ are the weighted least squares residuals given by

$$\hat{e}_i = \frac{y_i}{\pi_i} - \hat{B} ,$$

with

$$\hat{B} = \left(\sum_{j \in \mathscr{S}} c_j \right)^{-1} \sum_{i \in \mathscr{S}} c_i \, \frac{y_i}{\pi_i} .$$

The variance estimator (9) is clearly easier to implement than the Sen-Yates-Grundy estimator (Sen, 1953; Yates and Grundy, 1953) defined by (13). Moreover, the estimator (9) always gives positive estimates for the variance. Brewer (2002, Chap. 9) proposed two alternative choices for the c_i 's. A review of all the different choices for the c_i 's are given in Berger and Tillé (2011 p. 49).

Hájek (1964) showed that the estimator (9) is consistent under the maximum entropy sampling design. However, it is recommended to use the simple estimator (9) even if the sampling design is not the maximum entropy design (Deville, 1999).

As $\hat{\theta}(\mathscr{S}) = \hat{\text{var}}(\hat{Y})_H$ is consistent under $r(s)$, we can used Section 5's results, to derive regularity conditions under which (9) is consistent under any sampling designs which are different from the maximum entropy design. The following corollary follows from Theorem 1 and Corollary 1.

Corollary 3:

The Hájek (1964) *estimator* (9) *is consistent under a sampling design* $p(s)$ *if* (5) *or* (6) *or* (7) *holds.*

Berger (2005a) showed that under a set of regularity conditions, convergence (5) holds under Chao sampling. Hence the variance estimator (9) is consistent under Chao sampling.

Corollary 2 implies the following corollary.

Corollary 4:

The Hájek (1964) *estimator* (9) *is consistent under a sampling design* $p(s)$ *if* $p_t(\mathscr{S}) / r_t(\mathscr{S})$ *is uniformly bounded and if* (8) *holds.*

Note that Corollary 4 and Section 6's result imply that under Condition 1 and 2, Hájek (1964) estimator (9) is consistent under Rao-Sampford sampling.

It follows from corollary 3 or 4 that if $\hat{Y}$ is asymptotically normal, by Slutsky's lemma, we have that $(\hat{Y} - Y)\,\hat{\text{var}}(\hat{Y})_H^{-1/2}$ converges in distribution to a normal $N(0,1)$. This implies that confidence intervals based on $\hat{\text{var}}(\hat{Y})_H$ are asymptotically valid. The key requirement is that the π-estimator is asymptotically normal. Sufficient conditions for asymptotic normality have been investigated to a limited extent in the survey sampling literature, but some examples of conditions are given by Hájek (1964) and Rosén (1972). Hájek (1964) showed that $\hat{Y}$ is asymptotically normal under maximum entropy sampling. Berger (1998b) showed that $\hat{Y}$ is asymptotically normal when (6) holds. Hence the modes of convergence defined in Section 4 are also crucial for the asymptotic normality of $\hat{Y}$.

8. STRATIFICATION

Let $U_1, \cdots, U_H$ denote H strata, where $\bigcup_{h=1}^{H} U_h = U$. The size of U_h is denoted by N_h, and $\sum_{h=1}^{H} N_h = N$. Suppose that a sample s_h of fixed size n_h is selected without replacement with unequal probabilities within U_h. The random variable sample of stratum U_h is denoted by $\mathscr{S}_h$. The complete sample is $s = \bigcup_{h=1}^{H} s_h$, and the size of s is $n = \sum_{h=1}^{H} n_h$.

The Hájek stratified variance estimator is still given by (9), with

$$c_i = n_h (n_h - 1)^{-1} (1 - \pi_i) \quad (i \in U_h) \tag{10}$$

and with the following weighted least squares residuals

$$\hat{e}_i = \frac{y_i}{\pi_i} - \sum_{h=1}^{H} \hat{B}_h \, z_{ih} \, , \tag{11}$$

with

$$\hat{B}_h = \left(\sum_{j\in s} c_j \; z_{jh}^2 \right)^{-1} \sum_{i\in s} c_i \; z_{ih} \frac{y_i}{\pi_i} \, ;$$

where z_{ih} are strata indicators; that is, $z_{ih} = 1$ if $i \in U_h$ and otherwise $z_{ih} = 0$.

Let $\{ \mathscr{S}_{t,h} \}$ be a sequence of random variables which denotes random samples selected from the sequence of nested strata $U_{t,h}$ of size $N_{t,h}$ by a sequence of uni-stage probability sampling designs $\{ p_{t,h}(s_h) \}$. Let $\{ r_{t,h}(s_h) \}$ be the sequence of maximum entropy design of stratum $U_{t,h}$. We consider that $\bigcup_{h=1}^{H_t} U_{h,t} = U_t$; where H_t is the number of strata. In this section, $p_t(s)$ denotes the stratified Rao-Sampford design and $r_t(s)$ denotes the maximum entropy stratified design. The maximum entropy stratified design can be implemented by using the maximum entropy design within each stratum. Both designs ($p_t(s)$ and $r_t(s)$) have the same first-order inclusion probabilities $\pi_{t,i}$.

As $p_t(s)$ and $r_t(s)$ are stratified, we have that

$$\frac{p_t(s)}{r_t(s)} = \prod_{h=1}^{H_t} \frac{p_{t,h}(s_h)}{r_{t,h}(s_h)} \, ; \tag{12}$$

Assuming that Conditions 4-6 below hold, it can be shown that $p_t(\mathscr{S}) / r_t(\mathscr{S})$ is uniformly bounded and that (8) holds

Condition 3:

We assume that the number of strata H_t is bounded.

Condition 4:

We assume that $n_{t,h} / \tilde{d}_{t,h} = O(1)$ where $\tilde{d}_{t,h} = \sum_{i\in U_{t,h}} p_{t,i}(1 - p_{t,i})$

Condition 5:

We assume that that there exist constant $\kappa_{m,h}$ and $\kappa_{M,h}$ such that $\kappa_{m,h} < \hat{d}_{t,h} d_{t,h}^{-1} < \kappa_{M,h}$.

Condition 6:

We assume that $\lim_{t\to\infty} d_{t,h} = \infty$; where $d_{t,h} = \sum_{i\in U_{t,h}} \pi_{t,i}(1 - \pi_{t,i})$.

Conditions 4-6 imply that $p_{t,h}(\mathscr{S}_h) / r_{t,h}(\mathscr{S}_h) \xrightarrow{r} 1$ in probability and that $p_{t,h}(\mathscr{S}_h) / r_{t,h}(\mathscr{S}_h)$ is uniformly bounded under Rao-Sampford sampling (see Section 6). Furthermore, under Condition 3, we have that (8) holds (see Lemma 5.1.2 in Fuller, 1996 p. 217). It follows from (12) and Condition 3 that there exists a constant M such that $| 1 - p_t(s) / r_t(s) | < M$ for all s_t. Thus, $p_t(\mathscr{S}) / r_t(\mathscr{S})$ is uniformly bounded.

Condition 6 implies that the within strata Hájek variance is consistent under maximum entropy sampling (Hájek, 1964). Therefore corollary 4 implies that the within Hájek variance estimator is consistent under Rao-Sampford sampling. Hence the overall variance estimator (9) with c_i given by (10) and $\hat{e}_i$ given by (11), is consistent under Rao-Sampford sampling because (9) is a sum over the strata of within strata consistent estimators and the number of strata is bounded (Condition 3).

9. SIMULATION STUDY

In this section, the variance estimator (9) is compared numerically with the Sen-Yates-Grundy estimators given by

$$\hat{\text{var}}(\hat{Y})_{SYG} = \frac{1}{2} \sum_{i \in \mathscr{S}} \sum_{j \in \mathscr{S}} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 . \tag{13}$$

We use a population frame given in Valliant et al. (2000) and available at the John Wiley world wide web site ftp://ftp.wiley.com/public/sci_tech_med/finite_populations. This population frame is extracted from the September 1976 Current Population Survey in the United States. We replicated this population frame five times to create an artificial population of $N = 2390$ individuals from which samples are selected. The population is stratified into $H = 3$ strata with stratum sizes 1050, 1060 and 280. A proportional allocation is used. The variable of interest (y) is the weekly wages. Analysis of variance tests show that the stratum means of the weekly wages varies significantly between strata.

Three different designs are used to select units within each stratum: the Rao-Sampford design, the Chao (1982) design and the randomised systematic design (Madow, 1949). Let z_i be the number of hours worked per week. The π_i are proportional to z_i, or equal to one, for units with a large value for the z_i. The usual method described in Särndal et al. (1992, p. 89) is used to compute the π_i.

For each simulation, 10 000 samples were selected and, for each variance estimator, we computed, as percentages, the empirical relative bias,

$$\text{RB}(\hat{\text{var}}) = 100 \frac{E(\hat{\text{var}}) - \text{MSE}(\hat{Y})}{\text{MSE}(\hat{Y})} \%$$

and the empirical relative root mean squared error, $\text{MSE}(\hat{\text{var}})$. The quantity $\text{MSE}(\hat{Y})$ is the empirical mean squared error of $\hat{Y}$. Tables 1 gives the empirical relative biases and ratio of mean squared errors for different sample sizes n and for the three designs considered.

Table 1 shows that both variance estimators are unbiased. The Hájek simple variance estimator is as accurate as the Sen-Yates-Grundy estimator for the Rao-Sampford design, the randomised systematic design and the Chao design. Although, the Sen-Yates-Grundy estimator is theoretically unbiased, we observe a slight positive empirical relative bias. Note that with the Chao design, we observe a smaller relative bias for the Hájek variance estimator. The empirical bias of (13) is probably due to outlying units in the data and the structure of the joint-inclusion probabilities. As far as the mean squared error is concerned, we observe a slightly smaller mean squared error for the Hájek variance for small sample sizes, and for the randomised systematic design and the Chao design. For larger sample size, the Hájek estimator is slightly more precise than the Sen-Yates-Grundy estimator under Rao-Sampford sampling.

Table 1: Simulation study

Designs	n	$\text{RB}\{\hat{\text{var}}(\hat{Y})_{SYG}\}$	$\text{RB}\{\hat{\text{var}}(\hat{Y})_{H}\}$	$\frac{\text{MSE}\{\hat{\text{var}}(\hat{Y})_{H}\}}{\text{MSE}\{\hat{\text{var}}(\hat{Y})_{SYG}\}}$
Rao-Sampford	72	-0.55%	-0.68%	1.00
	119	1.31	1.01	0.98
	168	-4.30	-4.63	0.99
Randomised Systematic	72	0.41%	-0.03%	0.99
	119	-0.84	-1.37	1.00
	168	1.63	1.80	1.04
Chao	72	5.69%	1.50%	0.99
	119	4.37	-0.38	1.07
	168	3.92	1.98	1.04

Empirical relative biases (%) *and ratio of empirical mean squared errors of variance estimators* (13) *and* (9) *for three different stratified unequal probability sampling designs.*

10. SUMMARY

This paper proposes several new modes of convergence of sampling design towards the maximum entropy sampling design. It is shown how these modes of convergence can be used to prove consistency of the Hájek variance estimator. It is also shown that all the different modes of convergence hold with the Rao-Sampford design under a set of regularity conditions. A small simulation study is presented to support the theoretical findings in the paper.

ACKNOWLEDGEMENTS

I am grateful to an anonymous referee for helpful comments on a previous version of this paper.

REFERENCES

1. Aires, N. (1999). Algorithms to find exact inclusion probabilities for conditional Poisson sampling and Pareto πps sampling designs. *Methodology and Computing in Applied Probability*, 4, 457-469.
2. Aires, N. (2000). Comparisons between conditional Poisson sampling and Pareto πps sampling designs. *J. Statist. Plan. Infer.*, 82, 1-15.

3. Berger, Y.G. (1998a). Rate of Convergence to Asymptotic Variance for the Horvitz-Thompson Estimator. *J. Statist. Plan. Infer.*, 74, 149-168.
4. Berger, Y.G. (1998b). Rate of Convergence to Normal Distribution for the Horvitz-Thompson Estimator. *J. Statist. Plan. Infer.*, 67, 209-226.
5. Berger, Y.G. (2004). A Simple Variance Estimator for Unequal Probability Sampling Without Replacement. *J. Appl. Statist.*, 31, 305-315.
6. Berger, Y.G. (2005a). Variance estimation with Chao's sampling scheme. *J. Statist. Plan. Infer.*, 127, 253-77.
7. Berger, Y.G. (2005b). Inference with Highly Stratified Sampling Designs Having Few Units per Stratum. *Aust. & N. Z. J. Statist.*, 47, 365-373.
8. Berger, Y.G., and Tillé, Y. (2009), Sampling with Unequal Probabilities. In *Hanbook of Statistics,* Volume 29A: D. Pfeffermann and C.R. Rao. (editors). Elsevier.
9. Brewer, K.R.W. (2002). *Combined Survey Sampling Inference, Weighing Basu's Elephants*. London: Arnold.
10. Brewer, K.R.W. and Donadio, M.E. (2003). The high entropy variance of the Horvitz-Thompson estimator. *Survey Methodol*. 29, 189-196.
11. Brewer, K.R.W. and Hanif, M. (1983). *Sampling with Unequal Probabilities*. New York: Spinger-Verlag.
12. Chao, M.T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, 69, 653-56.
13. Chen, X.H., Dempster, A.P. and Liu, J.S. (1994). Weighting finite population sampling to maximise entropy. *Biometrika*, 81, 457-69.
14. Deville, J.C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques, *Survey Methodol*., 25, 193-203.
15. Fuller, W.A. (1996). *Introduction to Statistical Time Series*. New York: Wiley.
16. Grafström, A. (1964). Entropy of unequal probability sampling designs. *Statistical Methodology*, 7.
17. Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.*, 35, 1491-523.
18. Hájek, J. (1981). *Sampling in Finite Population*. New York: Marcel Dekker, Inc.
19. Hanurav, T.V. (1966). Some aspects of unified sampling theory. *Sankhya* A 28, 175-204.
20. Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.,* 47, 663-85.
21. Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc*., 377, 89-96.
22. Lehmann, E.L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag.
23. Madow, W.G. (1949). On the theory of systematic sampling II, *Ann. Math. Statist*., 20, 333-354.
24. Madow, L.H. and Madow, W.G. (1944). On the theory of systematic sampling, *Ann. Math. Statist*., 15, 1-24.
25. Narain, R.D. (1951). On Sampling without Replacement with Varying Probabilities, *J. Ind. Soc. Agri. Statist.*, 3, 169-174.
26. Rao, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Ind. Statist. Assoc*., 3, 173-80.
27. Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement, *I. Ann. Math. Statist*. 43, 373-97.

28. Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 494-513.
29. Särndal, C.E., Swenson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
30. Sen, P.K. (1953). On the Estimate of the Variance in Sampling with Varying Probabilities. *J. Indian Soc. Agric. Statist*., 5, 119-27.
31. Traat, I., Bondesson, L. and Meister, K. (2004). Sampling design and sample selection through distribution theory. *J. Statist. Plan. Infer*., 123, 395-413.
32. Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley.
33. Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc.*, B 1, 253-61.

APPENDIX A: Proof of Property 1

We have that

$$\lim_{t\to\infty}\sum_s r_t(s)\delta\{|u_t(s)-1|>\varepsilon\} = \lim_{t\to\infty} r_t\{|u_t(\mathscr{S})-1|>\varepsilon\} \qquad \text{(A1)}$$

where $u_t(s)=p_t(s)/r_t(s)$ and $u_t(\mathscr{S})=p_t(\mathscr{S})/r_t(\mathscr{S})$. By Markov's inequality

$$r_t\{|u_t(\mathscr{S})-1|>\varepsilon\}\le \varepsilon^{-1}E_{r_t}(|u_t(s)-1|), \qquad \text{(A2)}$$

where $E_{r_t}(\cdot)$ denotes the expectation with respect to the maximum entropy sampling design $r_t(s_t)$. As (6) holds, we have that $\lim_{t\to\infty} E_{r_t}(|u_t(s)-1|)=$ $\lim_{t\to\infty}\sum_s |p_t(s)-r_t(s)|=0$. Hence by combining (A1) and (A2), we have that for every $\varepsilon>0$, $\lim_{t\to\infty} r_t\{|u_t(\mathscr{S})-1|>\varepsilon\}=0$. This completes the proof.

APPENDIX B: Proof of Property 2

Using $|x-1|^2\le (x\log(x)-x+1)(4+2x)/3$, we have that

$$\| p_t - r_t \|_1 \le \sum_s r_t(s)\sqrt{(u_t(s)\log(u_t(s))-u_t(s)+1)(4+2u_t(s))/3}\ .$$

where $u_t(s)=p_t(s)/r_t(s)$. Using Cauchy's inequality, we have that

$$\begin{aligned} \| p_t - r_t \|_1 &\le \sqrt{\sum_s r_t(s)(u_t(s)\log(u_t(s))-u_t(s)+1)} \\ &\quad \sqrt{\sum_s (4+2u_t(s))r_t(s)/3} \\ &= \sqrt{2\sum_s p_t(s)\log(u_t(s))} \end{aligned} \qquad \text{(B1)}$$

Using (3), we have that

$$\begin{aligned} H(r_t) &= -\sum_s r_t(s)\log\left(\phi_t\ \delta\{\#s=n_t\}\prod_{i\in s_t}\alpha_i^t\right), \\ &= -\log(\phi_t)-\sum_s r_t(s)\delta\{\#s=n_t\}\sum_{i\in s}\log(\alpha_i^t) \\ &= -\log(\phi_t)-\sum_s r_t(s)\delta\{\#s=n_t\}\sum_{i\in U_t}\log(\alpha_i^t)\delta\{i\in s\} \\ &= -\log(\phi_t)-\sum_{i\in U_t}\log(\alpha_i^t)\pi_i^t \end{aligned} \qquad \text{(B2)}$$

where $\alpha_i^t = p_i^t/(1-p_i^t)$. Similarly, we also have that

$$-\sum_s p_t(s)\log(r_t(s)) = -\log(\phi_t) - \sum_{i\in U_t} \log(\alpha_i^t)\pi_i^t \tag{B3}$$

Combining equation (B2) and (B3), we have that

$$H(r_t) = -\sum_s p_t(s)\log(r_t(s)),$$

which implies

$$\begin{aligned}\sum_s p_t(s)\log(u_t(s)) &= \sum_s p_t(s)\log(p_t(s)) - \sum_s p_t(s)\log(r_t(s)) \\ &= H(r_t) - H(p_t) \\ &= D(p_t, r_t)\end{aligned} \tag{B4}$$

Combining (B1) and (B4), we have that $\| p_t - r_t \|_1 \leq \sqrt{2D(p_t, r_t)}$. This completes the proof.

APPENDIX C: Proof of Property 4

Using Property 1-3, it is only necessary to proof that (8) implies (7) and that (7) implies (5) when $p_t(\mathscr{S})/r_t(\mathscr{S})$ is uniformly bounded. By using Lehmann (1999, p. 53), it can be shown that (8) implies (7).Therefore, it is only necessary to show that (7) implies (5).

As $1+\log(x)-x \leq 0$, we have that

$$\sum_s p_t(s)(1+\log(u_t(s)) - u_t(s)) \leq 0,$$

where $u_t(s) = p_t(s)/r_t(s)$. As $\| p_t - r_t \|_2 \geq 0$, we have that

$$\sum_s p_t(s)(1+\log(u_t(s)) - u_t(s)) \leq \| p_t - r_t \|_2$$

Or equivalently

$$\sum_s p_t(s)\log(u_t(s)) \leq \| p_t - r_t \|_2 + \sum_s p_t(s)(u_t(s)-1)$$

Combining the last expression with (A6), we have that

$$D(p_t, r_t) \leq \| p_t - r_t \|_2 + \sum_s p_t(s)(u_t(s)-1). \tag{C1}$$

We also have that

$$\sum_s p_t(s)(u_t(s)-1) \leq \sum_{s_t} u_t(s) r_t(s) \,|\, u_t(s)-1 \,|. \tag{C2}$$

As $p_t(\mathscr{S})/r_t(\mathscr{S})$ is uniformly bounded, we have that there exists κ such that $u_t(s) \leq \kappa + 1$. Hence (C2) implies

$$\sum_s p_t(s)(u_t(s)-1) \le (\kappa+1)\sum_s r_t(s)\,|\,u_t(s)-1\,|\,. \tag{C3}$$

Using Cauchy's inequality and (C3), we have that

$$\sum_s p_t(s)(u_t(s)-1) \le (\kappa+1)\sqrt{\sum_s r_t(s)(u_t(s)-1)^2} = (\kappa+1)\sqrt{\| p_t - r_t \|_2}\,. \tag{C4}$$

Combining (C1) with (C4), we have that

$$D(p_t, r_t) \le \| p_t - r_t \|_2 + (\kappa+1)\sqrt{\| p_t - r_t \|_2}\,.$$

Hence (7) implies (5). This completes the proof.

APPENDIX D: Proof of Theorem 1

Let $\hat{\theta}_t(\mathscr{S})$ be a random variable which is consistent under the maximum entropy design; that is $\forall \varepsilon > 0$

$$\lim_{t\to\infty} r_t\{|\,\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)\,| > \varepsilon\} = 0\,. \tag{D1}$$

We have that

$$r_t\{|\,\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)\,| > \varepsilon\} = \sum_{s\in\Omega_t} r_t(s)\,;$$

where $\Omega_t = \{s_t : |\,\hat{\theta}_t(s) - \theta_t(U_t)\,| > \varepsilon\}$

Using the triangle inequality and the Cauchy's inequality, we have

$$\left|\sum_{s\in\Omega_t} p_t(s) - \sum_{s\in\Omega_t} r_t(s)\right| \le \sum_{s\in\Omega_t} |\,1-u_t(s)\,|\,r_t(s) \le \sum_s |\,1-u_t(s)\,|\,r_t(s) = \| p_t - r_t \|_1$$

where $u_t(s) = p_t(s)/r_t(s)$. Hence (6) implies

$$\lim_{t\to\infty}\left|\sum_{s\in\Omega_t} p_t(s) - \sum_{s\in\Omega_t} r_t(s)\right| = 0$$

Or equivalently,

$$\lim_{t\to\infty}\left| p_t\{|\,\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)\,| > \varepsilon\} - r_t\{|\,\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)\,| > \varepsilon\}\right| = 0 \tag{D2}$$

Equation (D1) and (D2) implies $\lim_{t\to\infty} p_t\{|\,\hat{\theta}_t(\mathscr{S}) - \theta_t(U_t)\,| > \varepsilon\} = 0$. The Theorem follows.

APPENDIX E

In this appendix we show that (8) holds, when $p(s)$ is the Rao-Sampford sampling design. By definition of the Rao-Sampford sampling design (e.g. Hájek, 1981 p86), we have

$$p_t(s) = \eta_t \; \hat{d}_t \; \delta\{\# s = n_t\} \prod_{i \in s} \frac{\pi_{t,i}}{1-\pi_{t,i}} , \tag{E1}$$

where $\hat{d}_t = \sum_{i\in s}(1-\pi_{t,i})$. The constant η_t is such that $\sum_s p_t(s) = 1$. Equation (E1) and (3) imply that for s such that $\# s = n_t$,

$$\frac{p_t(s)}{r_t(s)} = \frac{\eta_t}{\phi_t} \hat{d}_t \hat{Q}_t , \tag{E2}$$

where

$$\hat{Q}_t = \prod_{i\in s} \frac{\pi_{t,i}(1-p_{t,i})}{p_{t,i}(1-\pi_{t,i})} . \tag{E3}$$

As $\sum_s r_t(s) p_t(s) / r(s) = 1$, expression (E2) implies that

$$\frac{p_t(s)}{r_t(s)} = \frac{\hat{Q}_t \hat{d}_t}{\sum_s r_t(s) \hat{Q}_t \hat{d}_t} = \frac{\hat{Q}_t \hat{d}_t d_t^{-1}}{\tilde{Q}_t} ; \tag{E4}$$

where

$$\tilde{Q}_t = \sum_s r_t(s) \hat{Q}_t \hat{d}_t d_t^{-1} \tag{E5}$$

As our asymptotic framework assumes that $\lim_{t\to\infty} d_t = \infty$, we can use Theorem 5.1 in Hájek (1964) because the Rao-Sampford and the maximum entropy design have the same first-order inclusion probabilities given by $\pi_{t,i}$. Theorem 5.1 states that

$$\frac{\pi_{t,i}(1-p_{t,i})}{p_{t,i}(1-\pi_{t,i})} = 1 - \tau_{t,i} + o(\tilde{d}_t^{-1}) \tag{E6}$$

uniformly in $i \in U_t$, where $\tau_{t,i} = (\bar{\bar{p}}_t - p_{t,i}) / \tilde{d}_t$ and $\bar{\bar{p}}_t = \tilde{d}_t^{-1} \sum_{i\in U_t} p_{t,i}^2 (1-p_{t,i})$. Equation (E3) and (E6) imply that

$$\hat{Q}_t = \exp\left(\sum_{i\in s} \log(1 - \tau_{t,i} + o(\tilde{d}_t^{-1})) \right) . \tag{E7}$$

We have that

$$-\tilde{d}_t^{-1} < \tau_{t,i} < \tilde{d}_t^{-1} \tag{E8}$$

which implies that

$$\tau_{t,i} = O(\tilde{d}_t^{-1}) . \tag{E9}$$

Using the inequality $|\log(1-x)+x| \le x^2/(1-x)$, we have

$$\left|\sum_{i\in s}\log(1-\tau_{t,i}+o(\tilde{d}_t^{-1}))+\sum_{i\in s}\tau_{t,i}\right| \le \sum_{i\in s}\frac{(\tau_{t,i}+o(\tilde{d}_t^{-1}))^2}{1-\tau_{t,i}+o(\tilde{d}_t^{-1})} . \tag{E10}$$

As $\tau_{t,i} < 1/\tilde{d}_t$, we have $1-\tau_{t,i}+o(\tilde{d}_t^{-1}) > 1-1/\tilde{d}_t+o(\tilde{d}_t^{-1}) \ge M_3$, where M_3 is a strictly positive constant.

Hájek (1964, p. 1508) showed that $\tilde{d}_t \to \infty$, as $d_t \to \infty$. Furthermore, (E9), $n_t/\tilde{d}_t = O(1)$ and $\tilde{d}_t \to \infty$ imply that

$$\sum_{i\in s}\frac{(\tau_{t,i}+o(\tilde{d}_t^{-1}))^2}{1-\tau_{t,i}+o(\tilde{d}_t^{-1})} < \frac{1}{M_3}\sum_{i\in s}O(\tilde{d}_t^{-2}) = o_r(1) ;$$

where $o_r(1) \to 0$ in probability with respect to the maximum entropy design. Combining the last inequality with (E10), we have that $\forall \varepsilon > 0$

$$\lim_{t\to\infty} r_t\left\{\left|\sum_{i\in\mathscr{S}}\log(1-\tau_{t,i}+o(\tilde{d}_t^{-1}))+\sum_{i\in\mathscr{S}}\tau_{t,i}\right| > \varepsilon\right\} = 0 \tag{E11}$$

We also have

$$\sum_{i\in s}\tau_{t,i} = \hat{T}_t + \varphi_t , \tag{E12}$$

where

$$\hat{T}_t = \frac{1}{\tilde{d}_t}\left(\sum_{i\in U_t}p_{t,i}^2 - \sum_{i\in s}p_{t,i}\right),$$

$$\varphi_t = \frac{n_t}{\tilde{d}_t^2}\sum_{i\in U_t}(1-p_{t,i})(p_{t,i}^2 - p_i\tilde{p}_t)$$

and $\tilde{p}_t = n_t^{-1}\sum_{i\in U_t}p_{t,i}^2$. Applying Lemma 4.2 in Hájek (1964), we have $\lim_{t\to\infty} r_t\{|\hat{T}_t| > \varepsilon\} = 0$; that is, (E12) implies that $\forall \varepsilon > 0$

$$\lim_{t\to\infty} r_t\left\{\left|\sum_{i\in s}\tau_{t,i} - \varphi_t\right| > \varepsilon\right\} = 0 . \tag{E13}$$

Combining (E7) and (E11) and (E13), we have that $\forall \varepsilon > 0$

$$\lim_{t\to\infty} r_t\{|\log(\hat{Q}_t) - \log(\exp(-\varphi_t))| > \varepsilon\} = 0$$

which implies that $\forall \varepsilon > 0$

$$\lim_{t\to\infty} r_t\{|\hat{Q}_t - \exp(-\varphi_t)| > \varepsilon\} = 0 .$$

Furthermore, as $\lim_{t\to\infty} r_t\{|\hat{d}_t / d_t - 1| > \varepsilon\} = 0$ (see Hájek, 1964 p. 1503), we have

$$\lim_{t\to\infty} r_t\{|\hat{Q}_t\hat{d}_t d_t^{-1} - \exp(-\varphi_t)| > \varepsilon\} = 0 . \tag{E14}$$

Note that $\exp(-\varphi_t) > 0$, as it can be shown that $\varphi_t < n_t / \tilde{d}_t$ and $0 < n_t / \tilde{d}_t = O(1)$.

As $\log(1-y) \le -y$ when $y \le 1$, equation (E7) implies

$$\hat{Q}_t \le \exp\left(-\sum_{i\in s}(\tau_{t,i} + o(\tilde{d}_t^{-1}))\right) < \exp\left(\sum_{i\in s_h}\left(\frac{1}{\tilde{d}_t} + o(\tilde{d}_t^{-1})\right)\right), \tag{E15}$$

as $-\tau_{t,i} < 1/\tilde{d}_t$ (see (E8)) for all i. Inequality (E15) implies $\hat{Q}_t \le \exp(n_t / \tilde{d}_t(1+o(1)))$ which is bounded, as $n_t / \tilde{d}_t = O(1)$. As $\hat{Q}_t$ is bounded, $\hat{d}_t / d_t < \kappa_M$, $\varphi_t < n_t / \tilde{d}_t = O(1)$ and $\exp(-\varphi_t) > 0$, there exists a constant M_4 such that

$$\hat{Q}_t\hat{d}_t / d_t - \exp(-\varphi_t) \le M_4 . \tag{E16}$$

As $\log(1-x) \ge -x/(1-x)$, (E7) implies that

$$\hat{Q}_t \ge \exp\left(\sum_{i\in s}\frac{-\tau_{t,i} + o(\tilde{d}_t^{-1})}{1-\tau_{t,i} + o(\tilde{d}_t^{-1})}\right).$$

Using (E8), we have that

$$\hat{Q}_t > \exp\left(\sum_{i\in s}\frac{-\tilde{d}_t^{-1} + o(\tilde{d}_t^{-1})}{1-\tilde{d}_t^{-1} + o(\tilde{d}_t^{-1})}\right).$$

As there exists a constant M_5 such that $1-\tilde{d}_t^{-1} + o(\tilde{d}_t^{-1}) > M_5 > 0$, we have $\hat{Q}_t > \exp(-n_t / (\tilde{d}_t M_5)(1+o(1)))$. Thus, there is a constant M_6 such that $\hat{Q}_t > M_6$, as $n_t / \tilde{d}_t = O(1)$. Furthermore, as $\hat{d}_t / d_t > \kappa_m$ and $\varphi_t \ge 0$, there exists a constant M_7 such that

$$\hat{Q}_t\hat{d}_t d_t^{-1} - \exp(-\varphi_t) \ge M_7 . \tag{E17}$$

Now inequalities (E16) and (E17) show that $|\hat{Q}_t\hat{d}_t / d_t - \exp(-\varphi_t)|$ is uniformly bounded which implies that the expectation of $\hat{Q}_t\hat{d}_t / d_t - \exp(-\varphi_t)$ (with respect to the maximum entropy design) tends to zero as (E14) holds (see Lehmann, 1999 p53).; that is, $\lim_{t\to\infty}(\sum_s r_t(s)\hat{Q}_t\hat{d}_t d_t^{-1} - \exp(-\varphi_t)) = 0$ or equivalently

$$\lim_{t\to\infty}(\tilde{Q}_t - \exp(-\varphi_t)) = 0\,, \tag{E18}$$

Combining (E4) and (E14), we have that

$$\lim_{t\to\infty} r_t\{|\,(p_t(\mathscr{S})r_t(\mathscr{S})^{-1}-1)\tilde{Q}_t + \tilde{Q}_t - \exp(-\varphi_t)\,| > \varepsilon\} = 0 \tag{E19}$$

Using (E18), equation (E19) gives

$$\lim_{t\to\infty} r_t\{|\,(p_t(\mathscr{S})r_t(\mathscr{S})^{-1}-1)\tilde{Q}_t\,| > \varepsilon\} = 0$$

which implies

$$\lim_{t\to\infty} r_t\{|\,(p_t(\mathscr{S})r_t(\mathscr{S})^{-1}-1)\,| > \tilde{\varepsilon}\} = 0\,; \tag{E20}$$

where $\tilde{\varepsilon} = \varepsilon\,|\tilde{Q}_t|^{-1}$. Note that $|\tilde{Q}_t|^{-1}$ is bounded because Q_t and $\hat{d}_t d_t^{-1}$ are bounded. Hence (E20) holds for all $\tilde{\varepsilon} > 0$ which means that (8) holds. This completes the proof.

APPENDIX F

In this appendix, we show that $p_t(\mathscr{S})/r_t(\mathscr{S})$ is uniformly bounded. Inequality (E17) and equation (E5) implies

$$\tilde{Q}_t > M_7 + \exp(-\varphi)\,. \tag{F1}$$

Inequality (F1) combined with (E4) and (E16) gives

$$\frac{p_t(s)}{r_t(s)} < \frac{M_4 + \exp(-\varphi_t)}{M_7 + \exp(-\varphi_t)} < \frac{M_4 + 1}{M_7 + \exp(-n_t/\tilde{d}_t)}\,, \tag{F2}$$

as $0 \le \varphi_t < n_t/\tilde{d}_t$. The inequalities (F2) and $n_t/\tilde{d}_t = O(1)$ imply that $p_t(\mathscr{S}_t)/r_t(\mathscr{S}_t)$ is uniformly bounded.

APPENDIX G

We have that $p_{t,i} = \pi\{1 + p_{t,i}/\pi + (\pi_{t,i}/\pi - 1) - \pi_{t,i}/\pi\} \quad \le \pi\{1 + (p_{t,i} - \pi_{t,i})/\pi)\}$ $\le \pi(1+\varepsilon_{t,i})$; where $\varepsilon_{t,i} = p_{t,i}/\pi_{t,i}\,|1 - \pi_{t,i}/p_{t,i}| \to 0$ uniformly (see Hájek, 1964 p1508). Thus, we have that $p_{t,\min} \le \pi(1+\varepsilon_{t,\min})$, where $p_{t,\min}$ is the smallest $p_{t,i}$ for all $i \in U_t$ and $\varepsilon_{t,\min} = \varepsilon_{t,i}$ where i is such that $p_{t,i} = p_{t,\min}$. We have $n_t/\tilde{d}_t \le$ $1/(1-p_{t,\min}) \le\ 1/\{1-\pi(1+\varepsilon_{t,\min})\}$ which is bounded, as $\pi \ne 1$ and $\varepsilon_{t,\min} \to 0$. Thus, $n_t/\tilde{d}_t = O(1)$.