

## Empirical and Constrained Empirical Bayes Variance Estimation Under A One Unit Per Stratum Sample Design

Sepideh Mosaferi\*

### Abstract

A single primary sampling unit (PSU) per stratum design is a popular design for estimating the parameter of interest. Although the point estimator of the design is unbiased and efficient, an unbiased variance estimator does not exist. A common method for the variance estimation of this design is based on collapsing or combining two adjacent strata, but the attained estimator of variance is not design-unbiased, and the bias increases as the population means of collapsed strata become more different. If in some situations an unbiased estimator of variance is needed, the 1 PSU per stratum design with collapsed stratum variance estimator cannot be a good choice, and some statisticians prefer a design in which 2 PSUs per stratum are selected. In this paper, we first compare a 1 PSU per stratum design to a 2 PSUs per stratum design. Then, we propose an empirical Bayes estimator for the variance of 1 PSU per stratum design. But the empirical Bayes estimator over-shrinks towards the prior mean. To protect against this, we investigate the potential of the constrained empirical Bayes estimator. Using a simulation study, we show that the empirical Bayes and constrained empirical Bayes methods outperform the classical collapsed stratum variance method in terms of empirical relative mean squared error.

**Key Words:** Collapsing strata, Constrained empirical Bayes estimator, Empirical Bayes estimator, One PSU per stratum design, Two PSUs per stratum design, Variance estimation.

### 1. Introduction

A design in which one primary sampling unit (PSU) is selected in each stratum is theoretically efficient for providing an unbiased estimator of a population parameter. However, estimation of the variability of the attained estimator is impossible without considering any implicit assumptions such as collapsing strata; such assumptions produce a design biased estimator of the variance. Some examples that use stratified multi-stage with 1 PSU per stratum design include the Current Population Survey (CPS) and the National Crime Victimization Survey in the United States.

Due to the lack of an unbiased variance estimator for the 1 PSU per stratum design, some survey statisticians would prefer to select 2 PSUs per stratum since variance estimators for simple estimators like the Horvitz-Thompson estimator are unbiased. Surveys such as the Survey of Income and Program Participation (SIPP) of the U.S. Census Bureau and the U.S. Department of Agriculture's National Resources Inventory use a multi-stage two-per-stratum design and a stratified two-stage area sampling design, respectively. For the design with 2 PSUs per stratum, an unbiased variance estimator for linear estimators exists without any implicit assumptions, but the 1 PSU per stratum design still has its own popularity because it allows deeper stratification.

The collapsed stratum method for variance estimation was first introduced by Hansen et al. (1953). This method usually causes an overestimation in the variance of estimator; therefore, Hansen et al. (1953) and Isaki (1983) used some auxiliary variables, which are well-correlated with the expected values of the mean in each stratum, to reduce the bias of variance estimator. In 1969, Hartley et al. proposed a method of grouping strata where each group contains 7 to 15 strata and then applied a linear regression of the group means

---

\*Joint Program in Survey Methodology, University of Maryland, College Park, MD 20742.

on one or more auxiliary variables for estimating the variance of 1 PSU per stratum design; regression residuals used to estimate the stratum variance components. This method needs further evaluation study on the performance before being used, and the bias of the variance estimator depends on how well the regression model fits.

The idea of stratum boundaries being chosen by a random process prior to the sample selection was proposed by Fuller (1970). Fuller's method is biased when the stratum boundaries are not randomized beforehand. In 1987, Rust and Kalton examined the effects of collapsing strata in pairs, triples, and larger groups on the quality of the variance estimator and found that a greater level of collapsing is desirable when a small sample of PSUs is selected. They mentioned some factors that might be considered in deciding on the extent of collapsing.

Under the assumption of Durbin's (1967) sampling scheme within collapsed strata, Shapiro and Bateman (1978) applied the Yates-Grundy (1953) variance estimator. This variance can be biased upward, but the bias is relatively smaller than the collapsed method, and the variance estimator is more stable; however, for their empirical example, the authors did not consider the same number of units in collapsed strata; this resulted in the collapsed stratum variance performing poorly compared to the situation of having the same number of units in each collapsed strata.

Mantel and Giroux (2009) proposed a new approach based on the components of variance from different stages of sampling. They studied the Canadian Health Measures Survey (CHMS), a three-stage sample design, in which the number of PSUs per stratum is very small. In the study, they assumed that a randomized PPS systematic (RPPSS) sampling design was used for the CHMS; this assumption might introduce an unknown bias into the variance estimation. They also could not calculate an uncollapsed variance estimate for the Atlantic, the stratum with 1 PSU.

Recently, Breidt et al. (2014) proposed a nonparametric alternative method that replaces a collapsed stratum estimator by kernel-weighted stratum neighborhoods and used deviations from a fitted mean function to estimate the variance. They applied their method to the U.S. Consumer Expenditure Survey to demonstrate the superiority of their method over the collapsed stratum variance estimator. The estimator that they used is a natural nonparametric extension of linear models proposed by Hartley et al. (1969) and Isaki (1983).

In fact, most of the alternative recommended methods for the collapsed stratum variance are based on the existence of some concomitant or auxiliary information; nevertheless, this kind of desirable auxiliary information might not be readily available for all of strata. So, finding an acceptable comparative variance estimator for the variance of 1 PSU per stratum design without using auxiliary information is needed to decrease the bias and mse.

In Section 2 of this paper, we systematically compare the 1 PSU per stratum design (*Design 1*) to the 2 PSUs per stratum design (*Design 2*) based on the actual variance of the point estimator of population mean via a simulation study. In Section 3, we compare two design strategies analytically and through a Monte Carlo simulation to find out whether Design 1 performs better than Design 2 or not with respect to the coverage probability of mean.

In Section 4, we propose two alternatives to the collapsed variance estimator for the 1 PSU per stratum design using an empirical Bayes and a constrained empirical Bayes approaches. The findings of comparisons among the empirical Bayes estimator, constrained empirical Bayes estimator, and classical collapsed stratum method for the variance of 1 PSU per stratum design based on a simulation study are summarized in Section 5. We defer all of the necessary proofs to the Appendix Section.

## 2. Comparison of Design 1 and Design 2

For simplicity of exposition, we consider a stratified design with  $H$  strata where stratum  $h$  ( $h = 1, \dots, H$ ) consists of  $N_h$  units. A sample of  $n_h$  units is selected from stratum  $h$  using simple random sampling without replacement. In many applications, these units could be primary stage units. In this paper, we concentrate on single-stage sampling, and we are interested in estimating the finite population mean  $\bar{Y} = \sum_{h=1}^H W_h \bar{Y}_h$ , where  $W_h = N_h/N_T$ ,  $N_T = \sum_{h=1}^H N_h$ , and  $\bar{Y}_h$  is the finite population mean of the  $h$ th stratum. The Horvitz-Thompson unbiased estimator of the finite population mean ( $\bar{Y}$ ) is given by  $\bar{y}_{st} = \sum_{h=1}^H W_h \bar{y}_h$ , where  $\bar{y}_h$  is the sample mean for the  $h$ th stratum. The associated randomization-based variance is given by:

$$V(\bar{y}_{st}) = \sum_{h=1}^H W_h^2 \frac{1}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2, \tag{1}$$

where  $S_h^2 = \sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^2 / (N_h - 1)$  is the finite population variance for the  $h$ th stratum.

Here, we compare two popular design options:  $n_h = 1$  (one PSU per stratum design, say *Design 1*) and  $n_h = 2$  (two PSUs per stratum design, say *Design 2*). These two options are widely used in the context of stratified cluster sampling and stratified multi-stage sampling designs. To make a fair comparison of Design 1 and Design 2, we consider  $2H$  strata for Design 1 and  $H$  groups, each with two strata, for Design 2; therefore, we have an equal number of units (PSUs) for both designs.

The relative efficiency of Design 1 relative to Design 2 can be measured by design effect  $deff = V_2(\bar{y}_{st}) / V_1(\bar{y}_{st})$ , where  $V_1$  and  $V_2$  are the randomization-based variances of the same point estimator  $\bar{y}_{st}$  mentioned in (1). If  $deff > 1$ , Design 1 is more efficient than Design 2. On the other hand, if  $deff < 1$ , Design 1 is less efficient than Design 2. If  $deff = 1$ , the two designs are equivalent.

To compare Design 1 and Design 2, we conduct a Monte Carlo simulation study. Following Hansen et al. (1983), we generate a finite population of size  $N_T = 20,000$  units by drawing a random sample of size 20,000 from a bivariate superpopulation characterized by two dimensional random vector  $(x, y)$ , where the variable  $x$  has a gamma distribution with shape 2 and scale 5,  $f(x) = .04x \exp(-x/5)$ , and the variable  $y$  (conditional on  $x$ ) has a gamma distribution with density function  $g(y; x) = (1/b^c \Gamma(c)) y^{c-1} \exp(-y/b)$ , where  $c = .04x^{-3/2}(8 + 5x)^2$  and  $b = 1.25 x^{3/2} (8 + 5x)^{-1}$ .

To compare the effects of the number of strata  $H$  on the relative efficiency, we consider  $H = 10, 50$ , and 100 strata, which are formed based on the quantiles of  $x$ . We display the results in Table 1. Overall, Design 1 performs better than Design 2 with considerable efficiency for small  $H$ . However, efficiency diminishes as the number of strata increases (see Table 1).

**Table 1:** Comparison of Design 1 and Design 2 Based on the Number of Strata

<i>Design1</i>			<i>Design2</i>			<i>Comparison</i>
$H$	$N_h$	$V_1(\bar{y}_{st})$	$H$	$N_h$	$V_2(\bar{y}_{st})$	$deff = \frac{V_2(\bar{y}_{st})}{V_1(\bar{y}_{st})}$
10	2000	0.2515	5	4000	0.2759	1.0969
50	400	0.0464	25	800	0.0469	1.0104
100	200	0.0229	50	400	0.0232	1.0109

To be able to assess the effects of differences in population means and population variances within the collapsed strata on the randomization-based variance (1), we considered  $H = 10$  strata and 5 strata for Designs 1 and 2, respectively, and employed some

changes to the generated population’s means and variances within and between groups (collapsed strata). We separately generated data for each stratum from the Normal distribution,  $N(\mu = \text{mean}(y), \sigma^2 = \text{var}(y))$  and considered different coefficients of k, 2k, 3k, 4k, 5k (k = 1, 2) to implant some changes in  $\mu$  and  $\sigma^2$  for groups (g) 1, 2, 3, 4, and 5, respectively. The populations used for generating the data based on the Normal distribution are given in Table 2, and the results of comparisons associated to the groups of Table 2 are in Table 3.

**Table 2:** Generated Populations from the Normal Distribution for the Comparison

Case Study	Group 1	Group 2	Group 3	Group 4	Group 5
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $s_{g1}^2 \approx s_{g2}^2$	$N(\mu + 1, \sigma^2)$	$N(\mu + 2, 2\sigma^2)$	$N(\mu + 3, 3\sigma^2)$	$N(\mu + 4, 4\sigma^2)$	$N(\mu + 5, 5\sigma^2)$
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $s_{g1}^2 \approx s_{g2}^2$	$N(\mu + 1, \sigma^2)$	$N(\mu + 2, 2\sigma^2)$	$N(\mu + 3, 3\sigma^2)$	$N(\mu + 4, 4\sigma^2)$	$N(\mu + 5, 5\sigma^2)$
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $s_{g1}^2 \neq s_{g2}^2$	$N(\mu + 1, \sigma^2)$	$N(\mu + 2, 2\sigma^2)$	$N(\mu + 3, 3\sigma^2)$	$N(\mu + 4, 4\sigma^2)$	$N(\mu + 5, 5\sigma^2)$
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $s_{g1}^2 \neq s_{g2}^2$	$N(\mu + 1, \sigma^2)$	$N(\mu + 2, 2\sigma^2)$	$N(\mu + 3, 3\sigma^2)$	$N(\mu + 4, 4\sigma^2)$	$N(\mu + 5, 5\sigma^2)$

According to Table 3, when population means within groups are different, Design 1 is more efficient than Design 2; on the other hand, when the population means within groups are similar, there is no preference between two designs. In addition, changes in the population variances within the groups do not show any conspicuous effects on the efficiency.

**Table 3:** Comparison of Design 1 and Design 2 Based on the Differences of Means and Variances

Case Study	Design1			Design2			Comparison $deff = \frac{V_2(\bar{y}_{st})}{V_1(\bar{y}_{st})}$
	H	N <sub>h</sub>	V <sub>1</sub> ( $\bar{y}_{st}$ )	H	N <sub>h</sub>	V <sub>2</sub> ( $\bar{y}_{st}$ )	
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $S_{g1}^2 \approx S_{g2}^2$	10	2000	1.6198	5	4000	1.6198	1.0000
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $S_{g1}^2 \approx S_{g2}^2$	10	2000	1.6197	5	4000	1.9021	1.1743
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $S_{g1}^2 \neq S_{g2}^2$	10	2000	2.4310	5	4000	2.7047	1.1126
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $S_{g1}^2 \neq S_{g2}^2$	10	2000	2.4476	5	4000	2.4476	1.0000

### 3. Variance Estimation in Design 1 and Design 2

#### 3.1 Theoretical Expressions

As in section 2, we assume that we have two strata in each of the  $h$ th group and let  $N_{gi}$  denote the population size for the  $i$ th stratum within the  $g$ th group. Let  $y_{gij}$  denote the value of the characteristic of interest for the  $j$  unit in the  $i$  stratum within the  $g$  group ( $g = 1, \dots, H$ ,  $i = 1, 2$ ,  $j = 1, \dots, N_{gi}$ ). For simplicity in exposition, we assume  $N_{gi} = N$  ( $\forall g = 1, \dots, H$ ,  $i = 1, 2$ ), so  $N_T = 2HN$ ,  $W_{gi} = N_{gi}/N_T = 1/2H$ . We define:

$\bar{Y}_{gi} = N^{-1} \sum_{j=1}^N y_{gij}$ , finite population mean for the  $i$ th stratum within the  $g$ th group,

$S_{gi}^2 = (N-1)^{-1} \sum_{j=1}^N (y_{gij} - \bar{Y}_{gi})^2$ , finite population variance for the  $i$ th stratum in the  $g$ th group,

$\mu_{r,gi} = (N-1)^{-1} \sum_{j=1}^N (y_{gij} - \bar{Y}_{gi})^r$ , finite population  $r$ th central moment ( $r \geq 1$ ). Note that  $\mu_{1,gi} = 0$  and  $\mu_{2,gi} = S_{gi}^2$ .

We also assume that the finite population correction (FPC) factor is negligible as the sample size is only 1 or 2 per stratum, and  $N$  is large.

The true variance based upon Design 1 is:

$$V(\bar{y}_{st}) = \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2. \quad (2)$$

where  $S_h^2 = \sum_{j=1}^N (y_{hj} - \bar{Y}_h)^2 / (N-1)$ . We can rewrite (2) as  $V(\bar{y}_{st}) = \frac{1}{4H^2} \sum_{g=1}^H (S_{g1}^2 + S_{g2}^2)$ . The collapsed strata variance estimator is given by:

$$v(\bar{y}_{st}) = \frac{1}{2H^2} \sum_{g=1}^H s_g^2, \quad (3)$$

where  $s_g^2 = \sum_{i=1}^2 (y_{gi} - \bar{y}_g)^2$ , and  $\bar{y}_g = (y_{g1} + y_{g2})/2$ . The method relies on the implicit assumption:  $\bar{Y}_{g1} = \bar{Y}_{g2} = \bar{Y}_g$ .

Estimator (3) is design-biased, and its bias with respect to Design 1 is given by:

$$Bias(v(\bar{y}_{st})) = \frac{1}{4H^2} \sum_{g=1}^H (\bar{Y}_{g1} - \bar{Y}_{g2})^2. \quad (4)$$

As it is clear from (4), the bias is not related to the population variances within the groups. Wolter computed the bias of population total given the original sampling design (Wolter, 2007, p.51). Equation (4) suggests the strategy of how we can group strata to reduce the bias of collapsed stratum variance by putting more similar strata in pairs with respect to the characteristic of interest to minimize the difference  $|\bar{Y}_{g1} - \bar{Y}_{g2}|$ .

In order to find out the mean squared error (MSE) of  $v(\bar{y}_{st})$ , the theoretical variance of variance is needed. By ignoring the FPC, the variance is:

$$\begin{aligned} \text{Var}(v(\bar{y}_{st})) = & \frac{1}{16H^4} \sum_{g=1}^H \{ \mu_{4,g1} + \mu_{4,g2} + 2S_{g1}^2 S_{g2}^2 + 4(\bar{Y}_{g1} - \bar{Y}_{g2})^2 (S_{g1}^2 + S_{g2}^2) \\ & - (S_{g1}^2 - S_{g2}^2)^2 + 4(\bar{Y}_{g1} - \bar{Y}_{g2})(\mu_{3,g1} - \mu_{3,g2}) \}. \end{aligned} \quad (5)$$

If  $\mu_{4,g1} = \mu_{4,g2} = \mu_{4,g}$ ,  $\mu_{3,g1} = \mu_{3,g2} = \mu_{3,g}$ , and  $S_{g1}^2 = S_{g2}^2 = S_g^2$ , then

$$\text{Var}(v(\bar{y}_{st})) = \frac{1}{8H^4} \sum_{g=1}^H \{\mu_{4,g} + (S_g^2)^2 + 4S_g^2(\bar{Y}_{g1} - \bar{Y}_{g2})^2\}.$$

Therefore, the *MSE* of  $v(\bar{y}_{st})$  under all of these equality assumptions is:

$$\begin{aligned} \text{MSE}(v(\bar{y}_{st})) &= \text{Var}(v(\bar{y}_{st})) + \{\text{Bias}(v(\bar{y}_{st}))\}^2 \\ &= \frac{1}{8H^4} \sum_{g=1}^H \{\mu_{4,g} + (S_g^2)^2 + 4S_g^2(\bar{Y}_{g1} - \bar{Y}_{g2})^2\} \\ &\quad + \frac{1}{16H^4} \sum_{g=1}^H (\bar{Y}_{g1} - \bar{Y}_{g2})^4. \end{aligned}$$

The *MSE* is inversely related to the number of strata  $H$ . As a result, according to the asymptotic properties we can expect that as the number of strata  $H$  increases, *MSE* decreases. For Design 2, we ignore the FPC and use the standard variance of stratified estimator, which is unbiased under the design.

### 3.2 Simulation Study

We performed a simulation experiment to investigate the differences between the two designs with respect to the empirical coverage probability (CP) and average length (AL) of a nominal 95% confidence interval (CI) for  $\bar{y}_{st}$  under the two designs. The population used for this sub-section is similar to the one used in Table 3. The sample designs are the random selection of 1 PSU and 2 PSUs without replacement in each stratum for Design 1 and Design 2, respectively.

The process of sample selection was repeated 10,000 times. For each replication, we obtain  $\bar{y}_{st}$  and the two-sided 95% confidence interval,  $\bar{y}_{st} \pm 1.96\sqrt{v(\bar{y}_{st})}$ . For the variance, we use standard unbiased variance estimate for the 2 PSUs per stratum design and the collapsed strata variance given by (3) for the 1 PSU per stratum design. Table 4 displays the empirical coverage probability and average length for both designs.

When the means of collapsed strata are similar, two designs perform almost identically (see Table 4). On the other hand, the AL and CP are greater when the population means of collapsed strata are different; this can again reflect the important effect of  $|\bar{Y}_{g1} - \bar{Y}_{g2}|$  in collapsing. Based on Table 4, we cannot say which design is better when  $|\bar{Y}_{g1} - \bar{Y}_{g2}| \neq 0$ , since the more variability in Design 1 might result in the greater CP. In addition, as the number of strata increases the CP for the both designs approaches the nominal coverage probability 0.95, but the results are not displayed in Table 4 to save space and to be consistent with other tables.

### 4. E.B. and C.E.B. estimators for the Variance of Design 1

Let  $s_g^2 = (y_{g1} - y_{g2})^2/2$  denote the collapsed strata variance for the  $g$ th group of strata. Here  $y_{gi}$  denotes the sampled observation from the  $i$ th stratum in the  $g$ th group. We assume  $s_g^2/S_g^2$  has a chi-squared distribution with 1 degree of freedom ( $\chi^2(1)$ ) and an inverse gamma prior  $IG(a,a)$  for  $S_g^2$ . The posterior distribution  $\pi(S_g^2|s_g^2)$  is

$$\pi(S_g^2|s_g^2) \propto f_{S_g^2}(s_g^2)\pi(S_g^2) \propto \frac{(s_g^2)^{-1/2}e^{-s_g^2/2S_g^2}}{(S_g^2)^{1/2}}(S_g^2)^{-a-1}e^{-a/S_g^2}.$$

This is an inverse gamma distribution with shape  $a + \frac{1}{2}$  and scale  $a + \frac{s_g^2}{2}$ ,  $IG(a + \frac{1}{2}, a + \frac{s_g^2}{2})$ .

**Table 4:** Empirical Results of Simulation Study for Comparison of Design 1 and Design 2

Case Study	Design1		Design2	
	AL	CP%	AL	CP%
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $S_{g1}^2 \approx S_{g2}^2$	4.7309	88.81	4.7168	88.48
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $S_{g1}^2 \approx S_{g2}^2$	5.4929	92.35	5.0995	87.82
$\bar{Y}_{g1} \neq \bar{Y}_{g2}$ $S_{g1}^2 \neq S_{g2}^2$	6.4273	90.90	6.0775	88.88
$\bar{Y}_{g1} \approx \bar{Y}_{g2}$ $S_{g1}^2 \neq S_{g2}^2$	5.7853	89.32	5.7591	88.17

Under the squared error loss function,  $L(S_g^2, \hat{\delta}_g^B) \equiv (S_g^2 - \hat{\delta}_g^B)^2$ , the optimal Bayes estimator of  $S_g^2$  is the expectation of  $S_g^2$  conditional on  $s_g^2$ ,  $\{E(S_g^2|s_g^2)\}$ , which is,

$$\hat{\delta}_g^B = \lambda s_g^2 + (1 - \lambda) \frac{a}{a - 1} = \frac{2a + s_g^2}{2a - 1},$$

where  $\lambda$  equals to  $(2a - 1)^{-1}$ .

To estimate parameter “ $a$ ” based on the method-of-moments, the marginal distribution of  $s_g^2$  is needed. The marginal distribution of  $s_g^2$  is the  $F$ -distribution with 1 and  $2a$  degrees of freedom or equivalently,  $\sqrt{s_g^2}$  follows the Student’s  $t$ -distribution with  $2a$  degrees of freedom. The theoretical second order moment based on the Student’s  $t$ -distribution is  $E((\sqrt{s_g^2})^2) \equiv a/(a - 1)$ , which is valid for  $a$  greater than 1, is replaced by the empirical mean of the collapsed strata variances,  $s^2 = \sum_{g=1}^H s_g^2/H$ . Therefore, the solution is  $\hat{a}_{MM} = s^2/(s^2 - 1)$ , which yields an empirical Bayes estimator:

$$\hat{\delta}_g^{EB} = \frac{2\hat{a}_{MM} + s_g^2}{2\hat{a}_{MM} - 1}. \tag{6}$$

By substituting  $\hat{\delta}_g^{EB}$  from expression (6) into (3), the optimal estimator for the variance of 1 PSU per stratum design is attained:

$$\tilde{v}(\bar{y}_{st}) = \frac{1}{2H^2} \sum_{g=1}^H \hat{\delta}_g^{EB}. \tag{7}$$

In Design1 since the sample size from each stratum is small, the direct estimator  $s_g^2$  is overdispersed; therefore, under the Bayesian model we can show that:

$$E\left\{\frac{1}{H - 1} \sum_{g=1}^H (s_g^2 - s^2)^2\right\} = \frac{a^2(2a - 1)}{(a - 1)^2(a - 2)} > \frac{a^2}{(a - 1)^2(a - 2)} = E\left\{\frac{1}{H - 1} \sum_{g=1}^H (S_g^2 - S^2)^2\right\},$$

where  $S^2 = \sum_{g=1}^H S_g^2/H$ . While the direct estimator  $s_g^2$  shows overdispersion, the Bayes estimator shows underdispersion:

$$\begin{aligned}
 E\left\{\frac{1}{H-1} \sum_{g=1}^H (S_g^2 - S^2)^2 | s_g^2\right\} &= \frac{1}{H-1} \sum_{g=1}^H E\{(S_g^2 - S^2)^2 | s_g^2\} \\
 &= \frac{1}{H-1} \sum_{g=1}^H \{V[S_g^2 - S^2 | s_g^2] + (E[S_g^2 - S^2 | s_g^2])^2\} \\
 &= \frac{1}{H-1} \sum_{g=1}^H \{V[S_g^2 - S^2 | s_g^2] + (\hat{\delta}_g^B - \delta^B)^2\} \\
 &= \frac{1}{H-1} \sum_{g=1}^H V[S_g^2 - S^2 | s_g^2] + \frac{1}{H-1} \sum_{g=1}^H (\hat{\delta}_g^B - \delta^B)^2 \\
 &> \frac{1}{H-1} \sum_{g=1}^H (\hat{\delta}_g^B - \delta^B)^2,
 \end{aligned}$$

where  $\hat{\delta}^B = \sum_{g=1}^H \hat{\delta}_g^B / H$ . Hence,  $E\{\sum_{g=1}^H (S_g^2 - S^2)^2\} > E\{\sum_{g=1}^H (\hat{\delta}_g^B - \delta^B)^2\}$ . These inequalities hold when we use empirical Bayes estimator as well (see Lahiri (1990)).

The problem of underdispersion for these set of Bayes estimators is related to the fact that the standard Bayes estimator and empirical Bayes estimator shrink towards the prior mean  $a/(a - 1)$  specifically when the sample size is small. A solution for this problem might be attaching more weight to the direct estimator. To do so, we can match the ensemble variance by minimizing the posterior expected squared error loss  $E\{\sum_{g=1}^H (S_g^2 - \sigma_g^2)^2 | s_g^2\}$  subject to the following constraints:

- i)  $\sigma^2 = \frac{1}{H} \sum_{g=1}^H \sigma_g^2 = \frac{1}{H} \sum_{g=1}^H \hat{\delta}_g^B = \hat{\delta}^B,$
- ii)  $\frac{1}{H-1} \sum_{g=1}^H (\sigma_g^2 - \sigma^2)^2 = E\{\frac{1}{H-1} \sum_{g=1}^H (S_g^2 - S^2)^2 | s_g^2\}.$

We can write the posterior expected squared error loss  $E\{\sum_{g=1}^H (S_g^2 - \sigma_g^2)^2 | s_g^2\}$  as:

$$E\left\{\sum_{g=1}^H (S_g^2 - \sigma_g^2)^2 | s_g^2\right\} = E\left\{\sum_{g=1}^H (S_g^2 - \hat{\delta}_g^B)^2\right\} + \sum_{g=1}^H (\hat{\delta}_g^B - \sigma_g^2)^2. \tag{8}$$

So, in order to minimize the posterior expected squared error loss, it is sufficient to just minimize the last term of (8), which is related to  $\sigma_g^2$ . Using Lagrange multipliers  $\lambda_1$  and  $\lambda_2$ , we minimize  $\sum_{g=1}^H (\hat{\delta}_g^B - \sigma_g^2)^2$  subject to the constraints  $\sum_{g=1}^H \sigma_g^2 = C_1$  and  $\sum_{g=1}^H (\sigma_g^2 - \sigma^2)^2 = C_2$  or  $\sum_{g=1}^H \sigma_g^4 = C_2 + C_1^2/H$ , which means we minimize the objective function

$$\Phi = \sum_{g=1}^H (\hat{\delta}_g^B - \sigma_g^2)^2 - \lambda_1 \left(\sum_{g=1}^H \sigma_g^2 - C_1\right) - \lambda_2 \left(\sum_{g=1}^H \sigma_g^4 - C_2 - \frac{C_1^2}{H}\right),$$

with respect to the  $\sigma_g^2$ 's. Therefore, we can get

$$\sigma_{g,opt}^2 = \frac{1}{1 - \lambda_2} (\hat{\delta}_g^B + \frac{\lambda_1}{2}). \tag{9}$$

By imposing the constraints on (9), we obtain

$$\lambda_1 = 2\left\{(1 - \lambda_2) \frac{C_1}{H} - \hat{\delta}^B\right\}, \quad \lambda_2 = 1 - \left\{\sum_{g=1}^H (\hat{\delta}_g^B - \hat{\delta}^B)^2 / C_2\right\}^{1/2}.$$



We can rewrite  $C_1$  and  $C_2$  as follows:

$$C_1 = H\hat{\delta}^B, \quad C_2 = E\left\{\sum_{g=1}^H (S_g^2 - S^2)^2 | s_g^2\right\}.$$

Now by substituting  $\lambda_1$  and  $\lambda_2$  into (9), we get the constrained Bayes estimator:

$$\hat{\delta}_g^{CB} = \sigma_{g,opt}^2 = \hat{\delta}^B + \left\{ \frac{E[\sum_{g=1}^H (S_g^2 - S^2)^2 | s_g^2]}{\sum_{g=1}^H (\hat{\delta}_g^B - \hat{\delta}^B)^2} \right\}^{1/2} (\hat{\delta}_g^B - \hat{\delta}^B).$$

And using some algebra, we get

$$\hat{\delta}_g^{CB} = \hat{\delta}^B + \left\{ 1 + \frac{\frac{1}{H} \sum_{g=1}^H V(S_g^2 | s_g^2)}{\frac{1}{H-1} \sum_{g=1}^H (\hat{\delta}_g^B - \hat{\delta}^B)^2} \right\}^{1/2} (\hat{\delta}_g^B - \hat{\delta}^B). \quad (10)$$

In expression (10),  $V(S_g^2 | s_g^2)$  is the posterior variance, the variance of  $IG(a + \frac{1}{2}, a + \frac{s_g^2}{2})$ , which is  $2(2a + s_g^2)^2 / (2a - 1)^2(2a - 3)$ . Therefore, after some algebra,  $\hat{\delta}_g^{CB}$  in (10) can be written as follows:

$$\hat{\delta}_g^{CB} = \frac{1}{2a - 1} \left\{ 2a + s^2 + (s_g^2 - s^2) \left[ 1 + \frac{8(H - 1)(a^2 + s^2 a + \frac{1}{4H} \sum_{g=1}^H (s_g^2)^2)}{(2a - 3) \sum_{g=1}^H (s_g^2 - s^2)^2} \right]^{1/2} \right\}. \quad (11)$$

Since the constrained empirical Bayes estimator is close to the constrained Bayes estimator and the empirical Bayes estimator is close to the Bayes estimator, all of the mentioned results can be applied to the empirical Bayes and constrained empirical Bayes estimators (for more details on the equivalency between Bayes estimator (or constrained Bayes estimator) and empirical Bayes estimator (or constrained empirical Bayes estimator), readers should consult Lahiri (1990), and Ghosh and Lahiri (1987)).

As a consequence, the constrained empirical Bayes  $\hat{\delta}_g^{CEB}$  can be obtained by substituting  $\hat{a}_{MM}$  into (11), which yields:

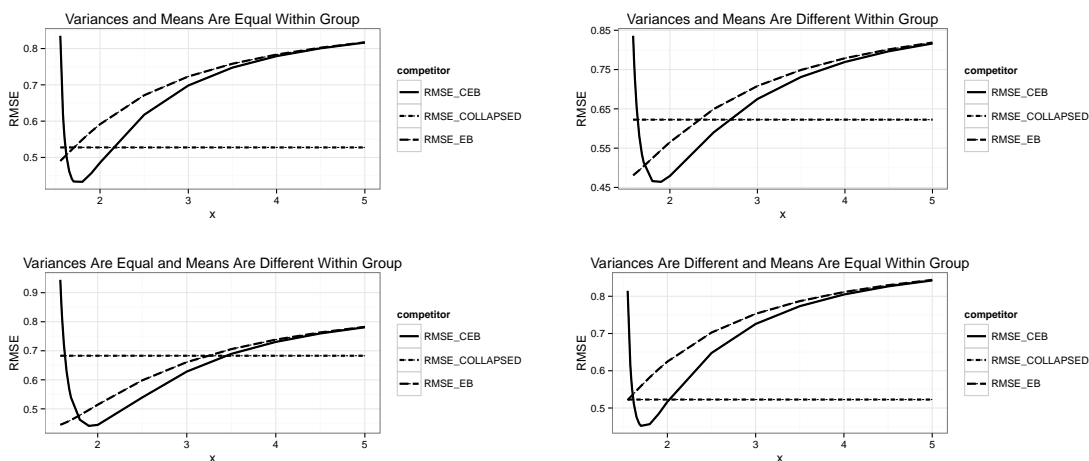
$$\hat{\delta}_g^{CEB} = \frac{1}{2\hat{a}_{MM} - 1} \left\{ 2\hat{a}_{MM} + s^2 + (s_g^2 - s^2) \times \left[ 1 + \frac{8(H - 1)(\hat{a}_{MM}^2 + s^2 \hat{a}_{MM} + \frac{1}{4H} \sum_{g=1}^H (s_g^2)^2)}{(2\hat{a}_{MM} - 3) \sum_{g=1}^H (s_g^2 - s^2)^2} \right]^{1/2} \right\}. \quad (12)$$

Since there is a possibility of receiving negative values for (12), which can be related to  $(s_g^2 - s^2)$ , we therefore use the  $\hat{\delta}_g^{EB}$  for the negative situations. The optimal estimator for the variance of 1 PSU per stratum design based on the constrained empirical Bayes is:

$$\tilde{v}_2(\bar{y}_{st}) = \frac{1}{2H^2} \sum_{g=1}^H \hat{\delta}_g^{CEB}. \quad (13)$$

## 5. Simulation Study and Results

For the sake of suitable comparisons among our candidate estimators (3), (7), and (13), we conduct a Monte Carlo study with 10,000 replications based on our proposed population in Table 3. The empirical relative mean squared error (RMSE) was found using the following formula:



**Figure 1:** Comparison Results of the Candidate Variances Based on the Empirical Relative MSE.

$$\sum_{r=1}^{10,000} \frac{RMSE_r}{10,000}; \quad RMSE_r = \frac{\sqrt{(v^*(\bar{y}_{st,r}) - V(\bar{Y}_{st}))^2}}{V(\bar{Y}_{st})},$$

where  $v^*(\bar{y}_{st,r})$  takes the values of our candidate estimators (3), (7), and (13), and  $V(\bar{Y}_{st})$  is the randomization-based variance in (1).

As we have different constraints for parameter  $a$ , theoretically and applicably, to make our estimators (6) and (12) become valid, we apply the intersection of constraints. So, the truncated  $\hat{a}_{MM}^* = \max(x, \hat{a}_{MM})$  is used, where  $x = 1.5 + e$ , and  $e$  can take values greater than zero. We consider different values for  $e$  to appropriately study the behaviors and/or effects of  $\hat{a}_{MM}^*$ , which is related to the lower bound ( $x$ ), on estimators (6), (12), and RMSEs as well. The results of comparisons are shown in Figure 1.

According to the graphs of Figure 1, the RMSEs of constrained empirical Bayes for all of the situations are greater than the competitors when  $x$  is really close to 1.5, since the denominator of (12) tends to zero for  $\hat{\delta}_g^{CEB}$ 's with  $x$  greater than  $\hat{a}_{MM}$ , and since the quantity of  $\hat{\delta}_g^{CEB}$ 's with  $x > \hat{a}_{MM}$  is reasonable; this can tremendously affect the results of RMSEs.

After moving away from this dangerous threshold (1.5), the RMSE of constrained empirical Bayes decreases compared to the RMSE of empirical Bayes. However, by increasing the value of  $x$  and assigning more weight  $(2a - 2)/(2a - 1)$  to the prior mean, the constrained empirical Bayes cannot perform well. Imposing great values for the constrain of  $a$  result in the underdispersion of empirical and constrained empirical Bayes estimators; therefore, their RMSEs will be increased. In addition, when the means of collapsed strata within the groups are different, empirical Bayes and constrained empirical Bayes outperform the classical collapsed stratum variance method.

### 6. Concluding Remarks

One PSU per stratum design has the advantage of deep stratification, which is efficient for estimating the finite population parameter of interest, but it is not possible to estimate the variance without making any implicit or explicit assumptions. The collapsed stratum variance estimator, a classical method for estimating the variance of the design, usually

suffers from the overestimation.

In this paper, for obtaining the exact MSE expression (except for the FPC) for the collapsed variance estimator, we assumed that a single element has been selected using a simple random sampling within each stratum. This assumption was made for the simplicity in exposition. The MSE for collapsed variance estimator to the case when the single PSU is selected using a general design can be extended in a straightforward way. The empirical Bayes and constrained empirical Bayes approaches developed in the paper are found to be promising alternatives to the traditional collapsed strata variance estimator for the 1 PSU per stratum design. The estimation of the prior parameter  $a$  in the EB and CEB approaches is a challenging problem. In order to ensure that the estimator of the prior parameter  $a$  is always within the admissible range, we proposed certain truncation strategies. In the future, we plan to explore a hierarchical Bayesian approach in an effort to rectify the problem.

### Acknowledgments

The author is grateful to her advisor, Professor Partha Lahiri, for suggesting the problem and his invaluable guidance and feedback throughout the research. The author also would like to thank Professor Richard L. Valliant for his comments which led to substantial improvement of the original manuscript.

### REFERENCES

- Breidt, F. J., Opsomer, J. D., and Borrego, I. S. (2014), "Nonparametric Variance Estimation under Fine Stratification: An Alternative to Collapsed Strata," *Journal of the American Statistical Association*, (Tentatively Accepted).
- Durbin, J. (1967), "Design of Multi-stage Surveys for the Estimation of Sampling Errors," *Applied Statistics*, 16, 152–164.
- Fuller, W. A. (1970), "Sampling with Random Stratum Boundaries," *Journal of the Royal Statistical Society B*, 32, 209–226.
- Ghosh, M., and Lahiri, P. (1987), "Robust Empirical Bayes Estimation of Means from Stratified Samples," *Journal of American Statistical Association*, 82, 1153–1162.
- Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory*, Vol. II, Wiley, New York.
- (1983), "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys," *Journal of the American Statistical Association*, 78, 776–793.
- Hartley, H. O., Rao, J. N. K., and Kiefer, G. (1969), "Variance Estimation with One Unit Per Stratum," *Journal of the American Statistical Association*, 64, 841–851.
- Isaki, C. T. (1983), "Variance Estimation Using Auxiliary Information," *Journal of the American Statistical Association*, 78, 117–123.
- Lahiri, P. (1990), "Adjusted Bayes and Empirical Bayes Estimation in Finite Population Sampling," *Sankhya: The Indian Journal of Statistics, Series B*, 52, 50–66.
- Mantel, H. and Giroux, S. (2009), "Variance Estimation in Complex Surveys with One PSU per Stratum," *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3069–3082.
- Rust, K. and Kalton, G. (1987), "Strategies for Collapsing Strata for Variance Estimation," *Journal of Official Statistics*, 3, 69–81.
- Shapiro, G. M. and Bateman, D. V. (1978), "A Better Alternative to the Collapsed Stratum Variance Estimate," *Proceedings of the Survey Research Methods Section, American Statistical Association*, 451–456.
- Wolter, K. M. (2007), *Introduction to Variance Estimation*, Springer Verlag.
- Yates, F. and Grundy, P. M. (1953), "Selection without Replacement from within Strata with Probabilities Proportional to Size," *Journal of the Royal Statistical Society B*, 15, 253–261.

## Appendix

### Proofs.

All of proofs in this section are based on the design-based application without considering any models.

*proof of (4):*

$$\begin{aligned}
 Bias(v(\bar{y}_{st})) &= E\left(\frac{1}{2H^2} \sum_{g=1}^H s_g^2\right) - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 = \frac{1}{2H^2} \sum_{g=1}^H E(s_g^2) - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 \\
 &= \frac{1}{2H^2} \sum_{g=1}^H E\left(\sum_{i=1}^2 (y_{gi} - \bar{y}_g)^2\right) - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 \\
 &= \frac{1}{2H^2} \sum_{g=1}^H E\left\{\frac{1}{2}(y_{g1} - y_{g2})^2\right\} - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 \\
 &= \frac{1}{4H^2} \sum_{g=1}^H E\{(y_{g1} - \bar{Y}_{g1}) - (y_{g2} - \bar{Y}_{g2}) + (\bar{Y}_{g1} - \bar{Y}_{g2})\}^2 \\
 &\quad - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 = \frac{1}{4H^2} \sum_{g=1}^H E\{(y_{g1} - \bar{Y}_{g1})^2 + (y_{g2} - \bar{Y}_{g2})^2 \\
 &\quad + (\bar{Y}_{g1} - \bar{Y}_{g2})^2 - 2(y_{g1} - \bar{Y}_{g1})(y_{g2} - \bar{Y}_{g2}) + 2(y_{g1} - \bar{Y}_{g1}) \\
 &\quad \times (\bar{Y}_{g1} - \bar{Y}_{g2}) - 2(y_{g2} - \bar{Y}_{g2})(\bar{Y}_{g1} - \bar{Y}_{g2})\} - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2.
 \end{aligned}$$

Under the stratified simple random sampling without replacement design, samples per stratum are selected independently; thus,  $y_{g1}$  and  $y_{g2}$  are independent. Also,  $\bar{Y}_{g1}$  and  $\bar{Y}_{g2}$ , the population means in each collapsed stratum of group  $g$  are fixed. As a result  $E(y_{g1} - \bar{Y}_{g1})(y_{g2} - \bar{Y}_{g2})$  equals to 0. Furthermore,  $E(y_{g1}) = \bar{Y}_{g1}$  and  $E(y_{g2}) = \bar{Y}_{g2}$ , so we can rewrite  $E(s_g^2)$  as follows:

$$E(s_g^2) = \frac{1}{2} E\{(y_{g1} - \bar{Y}_{g1})^2 + (y_{g2} - \bar{Y}_{g2})^2 + (\bar{Y}_{g1} - \bar{Y}_{g2})^2\}.$$

Thus;

$$\begin{aligned}
 Bias(v(\bar{y}_{st})) &= \frac{1}{4H^2} \sum_{g=1}^H E\{(y_{g1} - \bar{Y}_{g1})^2 + (y_{g2} - \bar{Y}_{g2})^2 + (\bar{Y}_{g1} - \bar{Y}_{g2})^2\} \\
 &\quad - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2 = \frac{1}{4H^2} \sum_{g=1}^H \{S_{g1}^2 + S_{g2}^2 + (\bar{Y}_{g1} - \bar{Y}_{g2})^2\} - \frac{1}{4H^2} \sum_{h=1}^{2H} S_h^2.
 \end{aligned}$$

And  $E(y_{g1} - \bar{Y}_{g1})^2 = \sum_{j=1}^N (Y_{g1j} - \bar{Y}_{g1})^2 / N = (1 - N^{-1})S_{g1}^2 \approx S_{g1}^2$ ,

$E(y_{g2} - \bar{Y}_{g2})^2 = \sum_{j=1}^N (Y_{g2j} - \bar{Y}_{g2})^2 / N = (1 - N^{-1})S_{g2}^2 \approx S_{g2}^2$ .

As  $\sum_{g=1}^H (S_{g1}^2 + S_{g2}^2) = \sum_{h=1}^{2H} S_h^2$ , the bias is:

$$Bias(v(\bar{y}_{st})) = \frac{1}{4H^2} \sum_{g=1}^H \{(\bar{Y}_{g1} - \bar{Y}_{g2})^2\}.$$

proof of (5):

$$\begin{aligned} Var(v(\bar{y}_{st})) &= Var\left(\frac{1}{2H^2} \sum_{g=1}^H s_g^2\right) = \frac{1}{4H^4} \sum_{g=1}^H Var(s_g^2) = \frac{1}{4H^4} \sum_{g=1}^H Var\left(\sum_{i=1}^2 (y_{gi} - \bar{y}_g)^2\right) \\ &= \frac{1}{4H^4} \sum_{g=1}^H Var\left(\frac{1}{2}(y_{g1} - y_{g2})^2\right) \\ &= \frac{1}{16H^4} \sum_{g=1}^H \{E(y_{g1} - y_{g2})^4 - \{E(y_{g1} - y_{g2})^2\}^2\}, \end{aligned}$$

where

$$\begin{aligned} E(y_{g1} - y_{g2})^4 &= E\{(y_{g1} - \bar{Y}_{g1})^4 + (y_{g2} - \bar{Y}_{g2})^4 + (\bar{Y}_{g1} - \bar{Y}_{g2})^4 \\ &\quad + 6(y_{g1} - \bar{Y}_{g1})^2(y_{g2} - \bar{Y}_{g2})^2 + 6(y_{g1} - \bar{Y}_{g1})^2(\bar{Y}_{g1} - \bar{Y}_{g2})^2 \\ &\quad + 6(y_{g2} - \bar{Y}_{g2})^2(\bar{Y}_{g1} - \bar{Y}_{g2})^2 + 4(y_{g1} - \bar{Y}_{g1})^3(\bar{Y}_{g1} - \bar{Y}_{g2}) \\ &\quad - 4(y_{g2} - \bar{Y}_{g2})^3(\bar{Y}_{g1} - \bar{Y}_{g2})\} = \mu_{4,g1} + \mu_{4,g2} + (\bar{Y}_{g1} - \bar{Y}_{g2})^4 \\ &\quad + 6S_{g1}^2 S_{g2}^2 + 6S_{g1}^2 (\bar{Y}_{g1} - \bar{Y}_{g2})^2 + 6S_{g2}^2 (\bar{Y}_{g1} - \bar{Y}_{g2})^2 \\ &\quad + 4\mu_{3,g1}(\bar{Y}_{g1} - \bar{Y}_{g2}) - 4\mu_{3,g2}(\bar{Y}_{g1} - \bar{Y}_{g2}), \end{aligned}$$

$$\text{and } \{E(y_{g1} - y_{g2})^2\}^2 = \{S_{g1}^2 + S_{g2}^2 + (\bar{Y}_{g1} - \bar{Y}_{g2})^2\}^2.$$

Therefore;

$$\begin{aligned} Var(v(\bar{y}_{st})) &= \frac{1}{16H^4} \sum_{g=1}^H \{\mu_{4,g1} + \mu_{4,g2} + (\bar{Y}_{g1} - \bar{Y}_{g2})^4 + 6S_{g1}^2 S_{g2}^2 \\ &\quad + 6S_{g1}^2 (\bar{Y}_{g1} - \bar{Y}_{g2})^2 + 6S_{g2}^2 (\bar{Y}_{g1} - \bar{Y}_{g2})^2 \\ &\quad + 4\mu_{3,g1}(\bar{Y}_{g1} - \bar{Y}_{g2}) - 4\mu_{3,g2}(\bar{Y}_{g1} - \bar{Y}_{g2}) \\ &\quad - \{S_{g1}^2 + S_{g2}^2 + (\bar{Y}_{g1} - \bar{Y}_{g2})^2\}^2\}. \end{aligned}$$

As a result,

$$\begin{aligned} Var(v(\bar{y}_{st})) &= \frac{1}{16H^4} \sum_{g=1}^H \{\mu_{4,g1} + \mu_{4,g2} + 2S_{g1}^2 S_{g2}^2 \\ &\quad + 4(\bar{Y}_{g1} - \bar{Y}_{g2})^2 (S_{g1}^2 + S_{g2}^2) \\ &\quad - (S_{g1}^2 - S_{g2}^2)^2 + 4(\bar{Y}_{g1} - \bar{Y}_{g2})(\mu_{3,g1} - \mu_{3,g2})\}, \end{aligned}$$

since

$$\begin{aligned} \mu_{3,g1} &= E(y_{g1} - \bar{Y}_{g1})^3 & , & & \mu_{3,g2} &= E(y_{g2} - \bar{Y}_{g2})^3 \\ \mu_{4,g1} &= E(y_{g1} - \bar{Y}_{g1})^4 & , & & \mu_{4,g2} &= E(y_{g2} - \bar{Y}_{g2})^4. \end{aligned}$$