Survey Methodology, June 1996
Vol. 22, No. 1, pp. 33–41
Statistics Canada

Small Area Estimation Under an Inverse Gaussian Model

Y.P. CHAUBEY, F. NEBEBE and P.S. CHEN[1]

ABSTRACT

In this paper, we consider analysis of variance methodology for inverse Gaussian distribution and adapt it for estimation of small area parameters in finite populations. It is demonstrated, through a Monte Carlo study, that these estimators offer a competitive choice for positively skewed survey data such as income or yield of a particular sector.

KEY WORDS: Interactions; Inverse Gaussian; Monte Carlo; Regression estimates; Synthetic estimates; Särndal-Hidiroglou estimator; Unbalanced model.

1. INTRODUCTION

Recently, a large number of methods appeared in the literature for the problem of small area estimation; for example Prasad and Rao (1990), Särndal and Hidiroglou (1989), Choudhry and Rao (1988), and Särndal (1984) and the references cited there, especially Särndal and Råbäck (1983), Fay and Herriot (1979), Schaible (1979), Holt, Smith and Tomberlin (1979), and Gonzalez and Hoza (1978), to name a few. The need for small area estimates of several characteristics of a given population has generated various useful procedures that produced realistic and sufficiently accurate estimates for local areas and other special subgroups. Several of the techniques suggested by the authors mentioned above were implicitly and/or explicitly model-based and utilized the standard normal theory. Others have tackled the provision of estimates for local areas from Bayesian and empirical Bayes perspectives by finding a compromise between the sample mean of an area (that is assumed to be normal) and an estimator based on regression on one or more covariates (see *e.g.*, Stroud 1987; MacGibbon and Tomberlin 1989). For an extensive review of recent developments in small area estimation, the reader may refer to Ghosh and Rao (1994).

The standard normal theory analysis of factorial experiments may be inappropriate to apply in situations where data are generated from markedly positively skewed distributions. While most of the inference procedures are analytically tractable, the accuracy and reliability of the results may be questionable in many practical applications. Thus, such an analysis based on positively skewed distributions is called for.

The objective of this paper is to consider inference procedures for unbalanced as well as balanced two-factor experiments under inverse Gaussian model that may be used to produce estimates for small regions. Hidiroglou and Särndal (1985) reported on a Monte Carlo study where a modified regression estimator is preferred as a compromise between the synthetic estimator and the generalized regression estimator. Särndal and Hidiroglou (1989) also presented further comparisons of estimators on the basis of conditional inference. The generalized regression estimator is basically derived from a super population regression model without any distributional assumptions. Chaubey (1991) considered super population models of Durbin (1959) with gamma auxiliary and inverse Gaussian auxiliary in which case the generalized regression estimator has the property of being the best linear unbiased predictor (see Prasad and Rao 1990). In fact, the best linear unbiased predictor for the population total does not depend on the form of the distribution of the characteristic variable, hence this technique is preferable given that maximum likelihood estimates (MLE) may be hard to obtain. As we have seen that the super population distributions (as transfused in the populations) may resemble closely to inverse Gaussian distributions for variety of populations we would like to exploit this aspect of the population.

The use of inverse Gaussian distribution is not merely a superficial one but it has been used successfully in many situations (see Folks and Chhikara 1978) and resembles closely to gamma, log normal and Weibull populations which are common in modeling positively skewed non negative random variables. In this paper, we study the use of inverse Gaussian model in applying to the small area estimation. The approach of Fries and Bhattacharyya (1983) which discusses the analysis of two factor experiments under an inverse Gaussian model is of major importance. The above paper gives estimation in balanced, no-interaction model. We have extended this approach to unbalanced case, which is essential for estimation of domain totals or means. In this respect the general multiple regression approach of Bhattacharyya and Fries (1986), and Whitmore (1983) may be adapted, but we have chosen to take the direct approach. In Section 2 we specify the

[1] Y.P. Chaubey, Professor, Department of Mathematics and Statistics; F. Nebebe, Associate Professor, Department of Decision Sciences & M.I.S.; and P.S. Chen, Research Assistant, Department of Finance, Concordia University, Montreal, Canada.

model and present our proposed estimators under the inverse Gaussian model. In Section 3, a numerical study is carried out for evaluation of the performance of the proposed estimator through Monte Carlo simulation. Finally, Section 4 presents summary and conslusions.

2. THE INVERSE GAUSSIAN REGRESSION MODEL FOR SMALL AREA ESTIMATION

Suppose that a finite population $\mathcal{U}$ is divided into D non-overlapping domains $U_{d.}$, $d = 1(1)D$, with $N_{d.}$ as the size of $U_{d.}$. The population is further divided along a second dimension, into G non-overlapping groups $U_{.g}$, $g = 1(1)G$, with the size of $U_{.g}$ denoted by $N_{.g}$. The cross-classification of domains and groups give rise to DG population cells U_{dg}, $d = 1(1)D$, $g = 1(1)G$, with N_{dg} as the size of U_{dg}. The population size N can then be expressed as $N = \sum_d N_{d.} = \sum_g N_{.g} = \sum_{dg} N_{dg}$. Our interest lies in estimating domain totals $t_d = \sum_{U_{d.}} y_k$, where y represents the characteristic variable and y_k is the observation on k-th unit. A sample s of size n is selected from $\mathcal{U}$ by a simple random sampling. Denote by $s_{d.}$, $s_{.g}$ and s_{dg} the parts of s that happen to fall in $U_{d.}$, $U_{.g}$ and U_{dg}. The corresponding sample sizes are denoted by $n_{d.}$, $n_{.g}$ and n_{dg}, respectively.

2.1 Regression Method for Inverse Gaussian Data

We refer readers to two recent comprehensive reviews about the developments in the inverse Gaussian distribution, namely, Chhikara and Folks (1989), and Iyengar and Patwardhan (1988). The probability density function of an inverse Gaussian variate with parameters (θ, σ), $IG(\theta, \sigma)$, is given by

$$f(y;\theta,\sigma) = (2\pi\sigma)^{-1/2} y^{-3/2} \exp[-(2\sigma y)^{-1}(y\theta^{-1} - 1)^2]; \tag{2.1}$$

with $y > 0, \theta > 0, \sigma > 0$. The mean and variance of this distribution are θ and $\theta^3\sigma$, respectively. Bhattacharyya and Fries (1982) proposed a reciprocal linear model for θ. Specifically, they assume a model of the form $\theta_k^{-1} = x_k'\eta$. An estimator of η, similar to the estimator of the regression parameter in the usual linear model (see Särndal 1984) in this situation is given by

$$\hat{\eta} = \left(\sum_{k \in S_d} \frac{x_k x_k' y_k}{\pi_k} \right)^{-1} \sum_{k \in S_d} \frac{x_k}{\pi_k}. \tag{2.2}$$

This is called pseudo Maximum Likelihood estimator, because it is obtained by unconditional maximization of the likelihood function and therefore $x_k'\hat{\eta} > 0$ may not be satisfied for all k. Then an estimator of the total t_d of the d-th domain in the spirit of Särndal's (1984) modified regression estimator may be constructed as

$$\hat{t}_{dIG} = \sum_{k \in U_{d.}} \hat{y}_k + \sum_{k \in S_{d.}} \frac{e_k}{\pi_k} \tag{2.3}$$

where $\hat{y}_k = x_k'\hat{\eta}$ and $e_k = y_k - \hat{y}_k$. In what follows, we denote the mean of the (d,g) cell by θ_{dg}, and consider the case of simple random sampling in which case π_k's are constant. We first discuss the prediction of observations for the use of (2.3) based on an additive effects model given by,

$$\theta_{dg}^{-1} = \mu + \alpha_d + \beta_g, \quad \sum \alpha_d = \sum \beta_g = 0, \tag{2.4}$$

where μ, α_d's and β_g's represent the overall effect, the domain or row effects, and the group or column effects, respectively. For the inverse Gaussian distribution we must also have $\theta_{dg} > 0$ for all (d,g) and $\sigma > 0$. Thus the parameters μ, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_D)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_G)$, and σ lie in the set $\Omega = \{(\mu, \alpha, \beta, \sigma) : \sum_d \alpha_d = 0, \sum_g \beta_g = 0; \mu + \alpha_d + \beta_g > 0, \forall(d,g); \sigma > 0\}$. Under this setup estimation of parameters for prediction can be accomplished through unconditional maximization of the likelihood function. Conditional on the population and the sample sizes n_{dg} and referring to (2.1) and (2.3), the log-likelihood function of the parameters is given by

$$\ell = -\frac{1}{2} \log \sigma \sum_d \sum_g n_{dg} - (2\sigma)^{-1} \sum_d \sum_g \sum_k y_{dgk}^{-1} [y_{dgk}(\mu + \alpha_d + \beta_g) - 1]^2. \tag{2.5}$$

We first note that the parameters are effectively given by $(\mu, \alpha_d, \beta_g, d = 1, 2, \ldots, D-1; g = 1, 2, \ldots, G-1)$. Thus, differentiating the above with respect to $(\mu, \alpha_d, \beta_g, d = 1, 2, \ldots, D-1; g = 1, 2, \ldots, G-1)$ and equating the resulting partial derivatives to zero gives the following equations for the estimators $(\hat{\mu}, \hat{\alpha}_d, \hat{\beta}_g, d = 1, 2, \ldots, D-1; g = 1, 2, \ldots, G-1)$,

$$\hat{\mu} y_{..} + \sum_{d=1}^{D-1} \hat{\alpha}_d (y_{d.} - y_{D.}) + \sum_{g=1}^{G-1} \hat{\beta}_g (y_{.g} - y_G) = n_{..},$$

$$\hat{\mu}(y_{d.} - y_{D.}) + \hat{\alpha}_d y_{d.} + \sum_{j=1}^{D-1} \alpha_j y_{D.} + \sum_{g=1}^{G-1} \hat{\beta}_g \{(y_{dg} - y_{Dg}) - (y_{dG} - y_{DG})\} = n_{d.} - n_{D.},$$

$$\hat{\mu}(y_{.g} - y_{.G}) + \sum_{d=1}^{D-1} \hat{\alpha}_d \{(y_{dg} - y_{dG}) - (y_{Dg} - y_{DG})\} + \hat{\beta}_g y_{.g} + \sum_{j=1}^{G-1} \hat{\beta}_j y_{.G} = n_{.g} - n_{.G}, \tag{2.6}$$

where the totals and means are represented by the notations

$$y_{dg} = \sum_k y_{dgk},\ y_{d.} = \sum_g y_{dg},\ y_{.g} = \sum_d y_{dg}, \quad (2.7a)$$

$$n_{d.} = \sum_g n_{dg},\ n_{.g} = \sum_d n_{dg},\ n_{..} = \sum_d \sum_g n_{dg}. \quad (2.7b)$$

The solutions $(\hat{\mu}, \hat{\alpha}_d, \hat{\beta}_g)$, $d = 1(1)D$, $g = 1(1)G$, provide the pseudo Maximum Likelihood estimator and may not yield nonnegative response estimates but will coincide with proper MLE as $n_{dg} \rightarrow \infty$ (see Fries and Bhattacharyya 1983) with probability one. Negative values of the response estimates may thus be truncated to zero.

In the case of the $IG(\theta, \sigma)$ model with interaction, the usual parameterization of the interaction effects suggests the model

$$\theta_{dg}^{-1} = \mu + \alpha_d + \beta_g + \gamma_{dg},$$

$$\sum \alpha_d = \sum \beta_g = \sum_d \gamma_{dg} = \sum_g \gamma_{dg} = 0, \quad (2.8)$$

where now γ_{dg} is the interaction effect when domain is at the d-th level and group is at the g-th level. The estimators of parameters may be obtained in this case following the method outlined above. However, noting that the maximum likelihood estimator (MLE) of θ_{dg} is $\bar{y}_{dg}$ and there is one to one relation between the parameters in the reparametrized model in terms of $(\mu, \alpha_d, \beta_g, \gamma_{dg})$ and the original parameters θ_{dg}, explicit formulae for the MLE of different parameters are not needed. Corresponding to equation (2.3), therefore, for a two-factor model with interaction, our estimator is

$$\hat{t}_{dWI} = \sum_g N_{dg} \bar{y}_{dg}, \quad (2.9)$$

which is the post stratified estimator and is not of further interest in small area estimation. For the model without interaction, the estimator is given as

$$\hat{t}_{dWOI} = \sum_g N_{dg} \hat{\theta}_{dg} + \sum_g \hat{N}_{dg} (\bar{y}_{dg} - \hat{\theta}_{dg}), \quad (2.10)$$

where $\hat{\theta}_{dg}^{-1} = \hat{\mu} + \hat{\alpha}_d + \hat{\beta}_g$, the estimators being obtained from (2.6) and $\hat{N}_{dg} = n_{dg} N / n_{..}$.

In order to judge the effectiveness of this estimator a numerical study has been performed and is reported in the following section.

3. A NUMERICAL STUDY OF THE INVERSE GAUSSIAN REGRESSION ESTIMATOR

In this section we provide the results of a simulation study which evaluates the performance of the estimators developed in the previous section. The modified regression estimator due to Särndal and Hidiroglou (1989) given below will be used as the bench mark for the above purpose;

$$\hat{t}_{dS-H} = \sum_g N_{dg} \bar{y}_{.g} + \sum_g F_d \hat{N}_{dg} (\bar{y}_{dg} - \bar{y}_{.g}), \quad (3.1)$$

where $F_d = N_{d.}/\hat{N}_{d.}$ if $\hat{N}_{d.} \geq N_{d.}$, otherwise $F_d = \hat{N}_{d.}/N_{d.}$. Here, $\hat{N}_{d.} = n_{d.} N / n_{..}$. An alternative form of this estimator which takes into account both group and domain effects can be obtained by replacing $\bar{y}_{.g}$ by $\bar{y}_{.g} + \bar{y}_{d.} - \bar{y}_{..}$ but this has not been pursued here. It should be noted that the above estimators cannot be computed when n_{dg} is zero. When this happens the estimators are simply taken to be the sample means of the respective domains. We also include the following modified version of $\hat{t}_{dWOI}$,

$$\hat{t}_{dWOIM} = \sum_g N_{dg} \hat{\theta}_{dg} + \sum_g F_d \hat{N}_{dg} (\bar{y}_{dg} - \hat{\theta}_{dg}), \quad (3.2)$$

for comparison.

3.1 Design of the Simulation Study

We consider Household Income data for Canadians in 1986, obtained from Household Income, Facilities and Equipment microdata tape of Statistics Canada (1987), for generating the values of parameters to be used for simulation. Using Household incomes, from these data, dividing them into 10 provinces and 6 educational groups, we first fit an inverse Gaussian model given by equation (2.4). The estimates of parameters are then used in forming the true parameters of the inverse Gaussian super population model which are summarized in appendix A. The values of D, G, N_{dg} are chosen from this population (see appendix B), where D represents the number of provinces (*i.e.*, $D = 10$) and G represents the number of education groups (*i.e.*, $G = 6$). Further sets of values of θ_{dg} and σ are obtained by considering various combinations of (c_1, c_2); $c_1 = 0(1)4$ and $c_2 = 1, .25, .1, .01$ where c_1 is used to transform θ_{dg} to $10^{-c_1}\theta_{dg}$ and c_2 is used to transform σ to $c_2\sigma$. Note that $c_1 = 0$ and $c_2 = 1$ gives the parameter values for the original population. Also, the higher values of c_1 indicate smaller values of the means and those of c_2 indicate higher value of the dispersion parameter.

For the simulation study, first we generate for a given set of θ_{dg} and σ values an inverse Gaussian random sample using the algorithm in Michael *et al.* (1976) with number of observations according to the values given in the appendix B. This random sample is then used as a finite population from which we select 1000 random samples for each of the sample fractions, 1%, and 5% with replacement. We had actually selected several random samples and obtained similar results as reported here. From each sample we computed the estimators of totals for the 10 domains using estimators $\hat{t}_{dS-H}$, $\hat{t}_{dWOI}$ and $\hat{t}_{dWOIM}$. The criteria for evaluating the performance of the estimators are the mean absolute relative error (MARE) and the absolute relative bias (ARB) defined as follows:

$$\mathrm{MARE}(\hat{t}_d) = \frac{1}{1000} \sum_{i=1}^{1000} \mid \hat{t}_{di} - t_d \mid / t_d \tag{3.3}$$

$$\mathrm{ARB}(\hat{t}_d) = \left| \frac{1}{1000} \sum_{i=1}^{1000} \hat{t}_{di} - t_d \right| / t_d . \tag{3.4}$$

Here $\hat{t}_d$ denotes a typical estimator of t_d and $\hat{t}_{di}$ denotes the value of the i-th Monte Carlo sample ($i = 1, \ldots, 1000$).

Table 1

Mean Absolute Relative Error (%) of Different Estimators

Domain	1% Sample			5% Sample			1% Sample			5% Sample		
	SH	WOI	WOIM	SH	WOI	WOIM	SH	WOI	WOIM	SH	WOI	WOIM
	$c_1 = 0, c_2 = 1$						$c_1 = 0, c_2 = .01$					
1	13.27	13.05	13.19	6.60	6.48	6.47	3.72	2.46	2.45	1.80	0.89	0.89
2	14.57	13.61	14.20	7.53	7.61	7.69	3.79	3.56	3.48	2.10	0.59	0.60
3	25.27	27.86	26.88	19.07	20.74	20.80	2.52	1.51	1.52	1.19	0.77	0.77
4	11.83	11.70	11.74	5.29	5.61	5.59	1.83	1.08	1.09	0.93	0.58	0.58
5	10.57	11.72	11.68	6.80	7.10	7.11	0.92	0.90	0.91	0.42	0.40	0.40
6	7.12	7.45	7.52	3.85	3.95	3.97	1.94	1.22	1.22	0.93	0.64	0.64
7	11.78	13.91	14.23	7.39	8.01	8.05	1.22	1.13	1.14	0.86	0.64	0.64
8	11.48	12.56	12.46	6.70	7.15	7.14	1.29	0.93	0.94	0.76	0.67	0.68
9	7.43	7.92	7.99	3.61	3.74	3.75	3.47	2.99	2.96	3.13	2.97	2.96
10	15.32	17.43	17.16	11.20	11.81	11.80	0.93	0.94	0.95	0.52	0.52	0.53
	$c_1 = 2, c_2 = 1$						$c_1 = 2, c_2 = .01$					
1	3.34	2.18	2.15	1.66	0.79	0.78	2.99	1.48	1.44	1.47	0.08	0.08
2	4.14	3.94	3.82	2.14	1.07	1.06	0.54	3.37	3.27	1.86	0.14	0.13
3	2.44	1.67	1.65	1.17	0.71	0.70	1.81	0.45	0.44	0.87	0.07	0.07
4	2.05	1.70	1.69	0.98	0.70	0.70	1.32	0.36	0.35	0.66	0.07	0.07
5	1.08	1.17	1.16	0.50	0.51	0.51	0.27	0.13	0.13	0.11	0.05	0.05
6	1.74	1.14	1.14	0.78	0.52	0.52	1.29	0.13	0.13	0.55	0.05	0.05
7	1.90	1.57	1.56	0.91	0.72	0.72	1.22	0.31	0.31	0.56	0.07	0.07
8	1.48	1.38	1.38	0.70	0.60	0.60	0.81	0.18	0.18	0.38	0.06	0.06
9	1.41	1.30	1.29	0.67	0.59	0.58	0.69	0.14	0.14	0.30	0.06	0.06
10	1.22	1.38	1.38	0.56	0.59	0.59	0.26	0.15	0.15	0.10	0.06	0.06
	$c_1 = 4, c_2 = 1$						$c_1 = 4, c_2 = .01$					
1	2.99	1.48	1.44	1.47	0.08	0.08	2.99	1.45	1.41	1.47	0.01	0.01
2	3.54	3.37	3.27	1.86	0.14	0.13	3.54	3.36	3.25	1.87	0.05	0.05
3	1.81	0.45	0.44	0.87	0.07	0.07	1.80	0.38	0.37	0.86	0.01	0.01
4	1.32	0.36	0.35	0.66	0.07	0.07	1.31	0.28	0.27	0.66	0.01	0.01
5	0.27	0.13	0.13	0.11	0.05	0.05	0.24	0.06	0.06	0.10	0.01	0.01
6	1.29	0.13	0.13	0.55	0.05	0.05	1.29	0.06	0.06	0.54	0.01	0.01
7	1.22	0.31	0.31	0.56	0.07	0.07	1.20	0.24	0.24	0.55	0.01	0.01
8	0.81	0.18	0.18	0.38	0.06	0.06	0.79	0.09	0.09	0.37	0.01	0.01
9	0.69	0.14	0.14	0.30	0.06	0.06	0.68	0.06	0.06	0.29	0.01	0.01
10	0.26	0.15	0.15	0.10	0.06	0.06	0.23	0.07	0.07	0.09	0.01	0.01

Table 2
Absolute Relative Bias (%) of Different Estimators

Domain	1% Sample			5% Sample			1% Sample			5% Sample		
	SH	WOI	WOIM	SH	WOI	WOIM	SH	WOI	WOIM	SH	WOI	WOIM
	$c_1 = 0, c_2 = 1$						$c_1 = 0, c_2 = .01$					
1	4.34	2.40	2.51	1.87	0.26	0.27	2.66	1.58	1.54	1.22	0.03	0.03
2	8.88	3.46	4.39	2.18	0.30	0.23	3.15	3.40	3.31	1.38	0.04	0.04
3	3.13	3.47	2.74	0.51	1.12	1.15	1.44	0.31	0.32	0.68	0.01	0.01
4	1.57	0.51	0.53	0.50	0.21	0.22	1.11	0.29	0.30	0.53	0.03	0.03
5	0.13	0.33	0.35	0.20	0.16	0.18	0.10	0.03	0.02	0.05	0.01	0.01
6	1.09	0.14	0.04	0.02	0.39	0.42	1.09	0.03	0.03	0.43	0.02	0.01
7	1.20	1.09	1.59	0.54	0.28	0.30	0.99	0.22	0.23	0.43	0.01	0.01
8	0.40	0.04	0.12	0.20	0.53	0.54	0.55	0.00	0.01	0.28	0.03	0.03
9	1.03	0.47	0.36	0.24	0.04	0.01	1.01	0.35	0.37	0.45	0.14	0.14
10	1.05	2.27	2.03	0.04	0.30	0.29	0.08	0.02	0.01	0.06	0.01	0.01
	$c_1 = 2, c_2 = 1$						$c_1 = 2, c_2 = .01$					
1	2.40	1.37	1.33	1.13	0.01	0.01	2.47	1.43	1.39	1.15	0.01	0.01
2	3.00	3.28	3.16	1.33	0.02	0.01	3.06	3.34	3.24	1.36	0.03	0.03
3	1.53	0.39	0.38	0.70	0.04	0.04	1.46	0.35	0.34	0.65	0.01	0.01
4	1.00	0.25	0.25	0.53	0.04	0.04	1.01	0.23	0.23	0.49	0.00	0.00
5	0.10	0.02	0.03	0.04	0.00	0.01	0.10	0.01	0.02	0.04	0.00	0.00
6	1.16	0.01	0.01	0.47	0.02	0.02	1.15	0.01	0.00	0.46	0.00	0.00
7	1.00	0.27	0.27	0.42	0.00	0.00	0.95	0.21	0.21	0.41	0.00	0.00
8	0.48	0.04	0.04	0.25	0.01	0.01	0.57	0.04	0.04	0.26	0.00	0.00
9	0.64	0.06	0.05	0.27	0.02	0.02	0.61	0.01	0.00	0.26	0.00	0.00
10	0.01	0.02	0.02	0.02	0.00	0.00	0.06	0.01	0.01	0.03	0.00	0.00
	$c_1 = 4, c_2 = 1$						$c_1 = 4, c_2 = .01$					
1	2.47	1.43	1.39	1.15	0.01	0.01	2.48	1.43	1.39	1.15	0.00	0.00
2	3.06	3.34	3.24	1.36	0.03	0.03	3.07	3.35	3.24	1.36	0.04	0.04
3	1.46	0.35	0.34	0.65	0.01	0.01	1.45	0.34	0.34	0.64	0.00	0.00
4	1.01	0.23	0.23	0.49	0.00	0.00	1.01	0.24	0.24	0.49	0.00	0.00
5	0.10	0.01	0.02	0.04	0.00	0.00	0.11	0.01	0.02	0.04	0.00	0.00
6	1.15	0.01	0.00	0.46	0.00	0.00	1.15	0.01	0.00	0.46	0.00	0.00
7	0.95	0.21	0.21	0.41	0.00	0.00	0.94	0.20	0.20	0.41	0.00	0.00
8	0.57	0.04	0.04	0.26	0.00	0.00	0.58	0.04	0.05	0.26	0.00	0.00
9	0.61	0.01	0.00	0.26	0.00	0.00	0.60	0.00	0.00	0.25	0.00	0.00
10	0.06	0.01	0.01	0.03	0.00	0.00	0.06	0.01	0.01	0.03	0.00	0.00

3.2 Analysis of Results

The MARE values computed according to (3.3) and the ARB values from (3.4) for the three estimators and for different sample sizes are reported in Tables 1 and 2, respectively for a selection of pairs (c_1, c_2). The values of c_1 are chosen to represent, large means (as in the original population, $c_1 = 0$), moderate means ($c_1 = 2$) and small means ($c_1 = 4$), whereas, the values chosen for c_2 represent the original dispersion parameter ($c_2 = 1$) and a further smaller value ($c_2 = .01$). It may be interesting to note that increasing c_1 by 1 while keeping c_2 fixed reduces the coefficient of variation by a factor of 10.

Some of the MARE and ARB values reported in Tables 1 and 2 are also plotted for visual inspection in Figures 1 and 2 for 1% samples, respectively.

When comparing the MARE and ARB values, reductions in biases as well as in relative errors are observed in many cases for both 1% and 5% samples. It is found that, the MARE and ARB values decrease with decreasing values of mean and dispersion parameter σ. Reductions

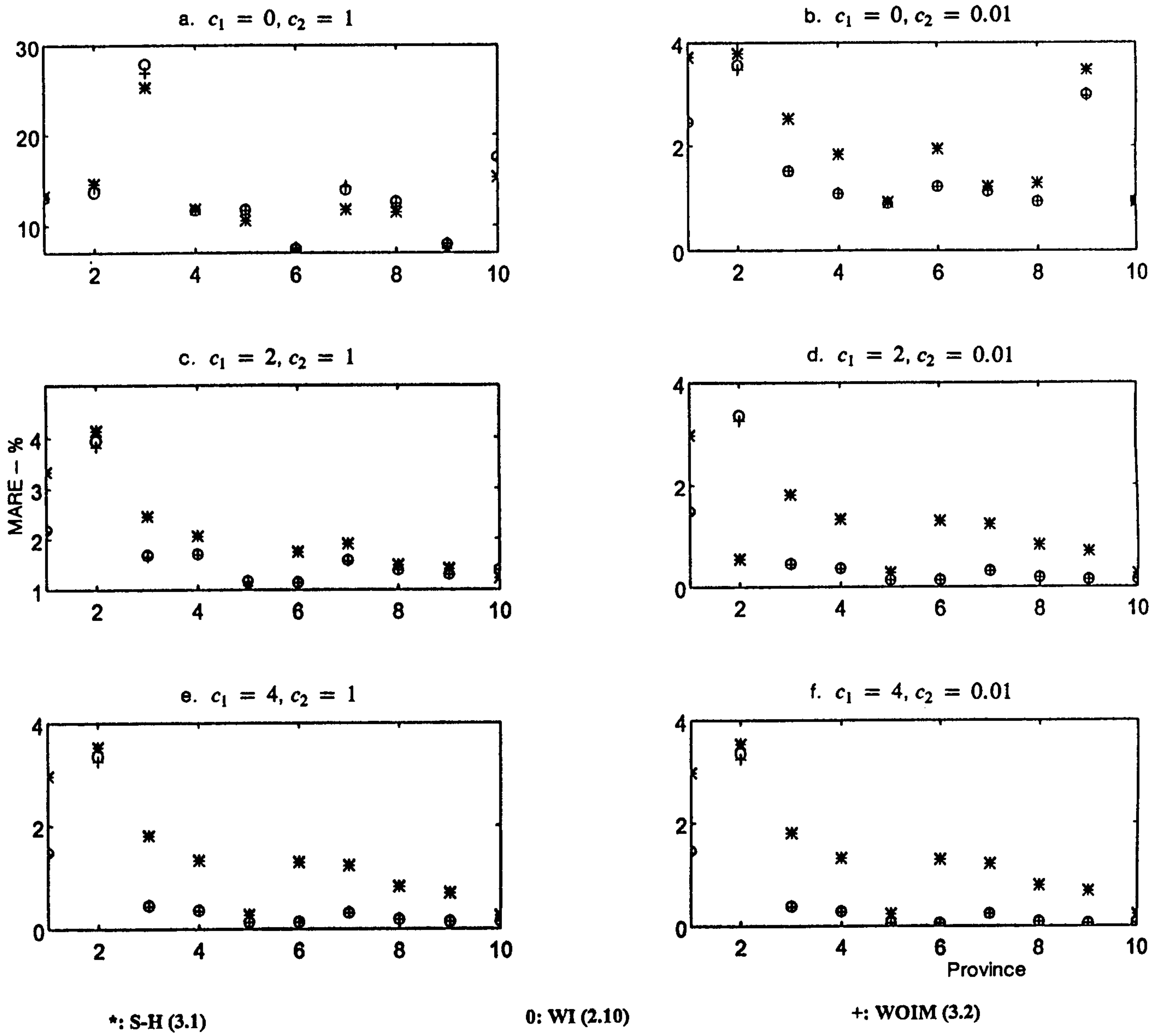

Figure 1. Mean absolute relative errors for different estimators for 1% sample.

are substantial, especially in case of 5% sample and/or when means are small. Note also that the reductions in bias are generally larger than reductions in the errors. We may note from Johnson and Kotz (1970, p. 141) that for fixed value of the mean, the standardized inverse Gaussian distribution tends to unit normal as the coefficient of variation tends to zero. Since larger gains in MARE and ARB values are noted for small values of the coefficient of variation, we conclude that proper modeling of the mean is important when the coefficient of variation is small for model based estimation.

We further find that $\hat{t}_{dWOI}$ and $\hat{t}_{dWOIM}$ have almost same MARE and ARB which indicates that the modification of the estimator in (2.10) is not necessary. It may be remarked that the estimator $\hat{t}_{dS-H}$, in contrast, has been demonstrated (see Hidiroglou and Särndal 1985) to be substantial improvement over the corresponding unmodified estimator due to Särndal (1984).

Owing to the criticism of $\hat{t}_{dWOI}$ and $\hat{t}_{dWOIM}$ as being model dependent, we want to defend these on the following grounds. The inverse Gaussian distribution offers a variety of shapes and may be able to approximate lognormal, gamma, Weibull and such other positively skewed shapes. If we suspect that the principal characteristic is positively skewed, then the methodology we discussed here is viable and useful.

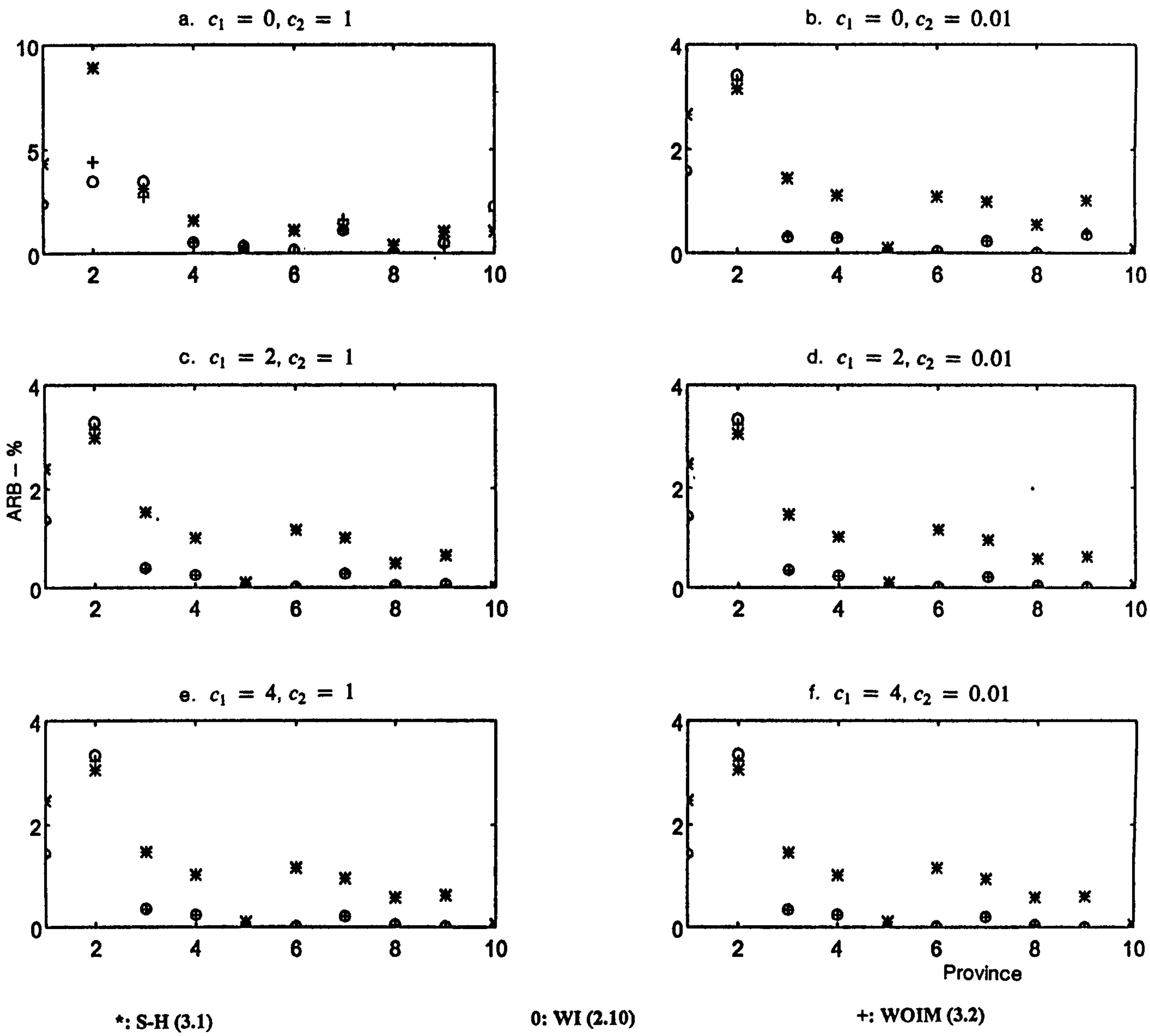

Figure 2. Absolute relative biases for different estimators for 1% sample.

4. SUMMARY AND CONCLUSIONS

The generalization of analysis of variance methodology for inverse Gaussian population for unbalanced design was considered. The models without interactions of factors were studied and applied to the problem of estimation of small area parameters in finite populations. Using Canadian survey data, synthetic populations were generated in a Monte Carlo study. Through this we demonstrated that the proposed estimators perform well under a variety of conditions when the population can be regarded as a random sample from some inverse Gaussian distribution. This approach offers a competitive choice for estimation of parameters in positively skewed survey data.

ACKNOWLEDGEMENTS

Support to Y.P. Chaubey and F. Nebebe by the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. The authors thank the Editor, Professors J.N.K. Rao and A.B. Sim and the referees for their helpful comments.

APPENDIX A
Values of the Parameters for Generation of the IG Population

$$\mu = 3.13241147 \times 10^{-5}, \quad \sigma = 2.5447984 \times 10^{-5}$$

d	1	2	3	4	5
$10^6 \times \alpha_d$	3.1902855	2.8235779	1.5676078	.8056079	−.95350458

d	6	7	8	9	10
$10^6 \times \alpha_d$	−4.0661125	.49944356	.0061694263	−2.7414128	−1.1316622

g	1	2	3	4	5	6
$10^5 \times \beta_g$	1.0938451	.36781639	−.012707035	−.11561414	−.30936835	−1.023972

θ_{dg} values:

d/g	1	2	3	4	5	6
1	22,000.82	26,183.11	29,080.48	29,977.59	31,826.13	41,195.19
2	22,179.76	26,436.94	29,393.94	30,310.79	32,201.96	41,827.05
3	22,815.33	27,344.90	30,520.70	31,510.37	33,559.25	44,146.20
4	23,219.00	27,926.81	31,247.41	32,285.58	34,439.96	45,682.95
5	24,207.76	29,369.63	33,064.91	34,229.61	36,661.02	49,674.90
6	26,180.44	32,324.63	36,858.30	38,311.45	41,383.33	58,760.34
7	23,385.24	28,167.65	31,549.24	32,607.90	34,806.97	46,330.96
8	23,658.15	28,564.53	32,047.98	33,140.96	35,415.03	47,414.57
9	25,302.90	30,997.31	35,142.43	36,461.01	39,232.58	54,516.76
10	24,312.62	29,524.12	33,260.85	34,439.64	36,902.04	50,118.45

APPENDIX B
Values of the Cell Sizes N_{dg}

d/g	1	2	3	4	5	6	Total
1	627	360	277	84	215	110	1,673
2	285	212	198	72	68	83	918
3	597	483	616	148	204	231	2,279
4	729	397	568	151	239	219	2,303
5	1,372	761	1,216	202	473	511	4,535
6	1,177	888	1,795	517	707	800	5,884
7	639	432	673	165	236	222	2,367
8	850	512	888	264	349	297	3,160
9	700	699	1,350	385	696	572	4,401
10	456	540	1,083	342	393	407	3,221

REFERENCES

BHATTACHARYYA, G.K., and FRIES, A. (1986). On the inverse Gaussian multiple regression and model checking procedures. In *Reliability and Quality Control*, (A.P. Basu, Ed.). New York: North Holland, 86-100.

CHAUBEY, Y.P. (1991). A study of ratio and product estimators under super population. *Communications in Statistics*, A, 20 (5 and 6), 1731-1746.

CHOUDHRY, G.H., and RAO, J.N.K. (1988). Evaluation of small area estimators: An empirical study. Paper presented at the International Symposium on Small Area Statistics, New Orleans.

CHHIKARA, R.S., and FOLKS, J.L. (1989). *The Inverse Gaussian Distribution*. New York: Marcel Dekker, Inc.

DURBIN, J. (1959). A Note on the application of Quenouille's method of bias reduction to the estimation of ratio. *Biometrika*, 46, 477-480.

FAY, R.E., and HERRIOT, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

FOLKS, J.L., and CHHIKARA, R.S. (1978). The inverse Gaussian distribution and its statistical applications-A review. *Journal of the Royal Statistical Society*, Series B, 40, 263-275.

FRIES, A., and BHATTACHARYYA, G.K. (1983). Analysis of two-factor experiments under an inverse Gaussian model. *Journal of the American Statistical Association*, 78, 820-826.

GONZALES, M.E., and HOZA, C. (1978). Small area estimation with application to unemployment and housing estimates. *Journal of the American Statistical Association*, 73, 7-15.

GHOSH, M., and RAO, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-76.

HIDIROGLOU, M.A., and SÄRNDAL, C.-E. (1985). An empirical study of some regression estimations for small domains. *Survey Methodology*, 6, 65-77.

HOLT, D., SMITH, T.M.F., and TOMBERLIN, T.J. (1979). A model-based approach to estimation for small subgroups of a population. *Journal of the American Statistical Association*, 74, 405-410.

IYENGAR, S., and PATWARDHAN, G. (1988). Recent developments in the inverse Gaussian distribution. In *Handbook of Statistics*, New York: Elsevier Science 479-490.

JOHNSON, N. L., and KOTZ, S. (1970). *Continuous Univariate Distributions-1, Distributions in Statistics*. New York: Wiley.

MACGIBBON, B., and TOMBERLIN, T.J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology*, 15, 237-252.

MICHAEL, J.R., SCHUCANY, W.R., and HASS, R.W. (1976). Generating random variables using transformations with multiple roots. *American Statistician*, 30(2), 88-90.

PRASAD, N.G.N., and RAO, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

PURCELL, N.J., and KISH, L. (1980). Postcensal estimates for local areas (or domains). *Bulletin of the International Statistical Institute*, 48, 3-18.

SÄRNDAL, C.-E. (1984). Design consistent versus model dependent estimation for small domains. *Journal of the American Statistical Association*, 79, 624-631.

SÄRNDAL, C.-E., and HIDIROGLOU, M.A. (1989). Small domain estimation: A conditional analysis. *Journal of the American Statistical Association*, 84, 266-275.

SÄRNDAL, C.-E., and RÅBÄCK, G. (1983). Variance reduction and unbiasedness for small domain estimators. *Statistical Review*, 21, (Essays in Honor of T.E. Dalenius), 33-40.

SCHAIBLE, W.L. (1979). A composite estimator for small area statistics. In *Synthetic Estimates for Small Areas*, NIDA Research Monograph 24, (J. Steinberg, Ed.). Rockville, MD: National Institute on Drug Abuse, 36-53.

STATISTICS CANADA (1987). Microdata file, Household Income, Facilities and Equipment (1987), Statistics Canada, Household Surveys Division.

STROUD, T.W.F. (1987). Bayes and empirical Bayes approaches to small area estimation. In *Small Area Statistics*, (R.Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh, Eds.). New York: Wiley, 124-137.

WHITMORE, G.A. (1983). A regression method for censored inverse Gaussian data. *The Canadian Journal of Statistics*, 11, 305-315.