3. Direct estimators for domain estimation

The HT type estimator does not incorporate auxiliary information. GREG estimation is assisted by a model fitted at the domain level and uses auxiliary data from the domain. Calibration incorporates auxiliary data from the domain of interest or from a higher-level aggregate. All these estimators are direct because the y-values are taken from the domain of interest. When domain membership is known for all population elements, domain sizes N_d are also known.

3.1. Horvitz–Thompson estimator

The basic design-based direct estimator of the domain total t_d is the HT estimator, also known as the Narain-Horvitz-Thompson (NHT) and the *expansion estimator*:

$$\hat{t}_{d\mathrm{HT}} = \sum_{k\in U_d} I_k y_k/\pi_k = \sum_{k\in s_d} y_k/\pi_k = \sum_{k\in s_d} a_k y_k \tag{1}$$

(Horvitz and Thompson, 1952; Narain, 1951; notation as in Section 2.1). HT estimates of domain totals are additive: they sum up to the HT estimator $\hat{t}_{\mathrm{HT}} = \sum_{k\in s} a_k y_k$ of the population total. As $E(I_k) = \pi_k$, the HT estimator is design unbiased for t_d. Under mild conditions on the π_k, the corresponding mean estimator $\hat{t}_{d\mathrm{HT}}/N_d$ is also design consistent (Isaki and Fuller, 1982). The estimator $\hat{t}_{d\mathrm{HT}}$ has design variance

$$\begin{aligned}\mathrm{Var}\big(\hat{t}_{d\mathrm{HT}}\big) &= E\left(\sum_{k\in U_d} \frac{I_k - \pi_k}{\pi_k} y_k\right)^2 = \sum_{k\in U_d}\sum_{l\in U_d} E(I_k-\pi_k)(I_l-\pi_l)\,\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} \\ &= \sum_{k\in U_d}\sum_{l\in U_d} (\pi_{kl}-\pi_k\pi_l)\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} = \sum_{k\in U_d}\sum_{l\in U_d} (a_k a_l/a_{kl} - 1) y_k y_l. \end{aligned}\tag{2}$$

From $a_{kl}E(I_k I_l) = 1$, we see that an unbiased estimator for the design variance is

$$\hat{V}\big(\hat{t}_{d\mathrm{HT}}\big) = \sum_{k\in U_d}\sum_{l\in U_d} a_{kl} I_k I_l (a_k a_l/a_{kl} - 1) y_k y_l = \sum_{k\in s_d}\sum_{l\in s_d} (a_k a_l - a_{kl}) y_k y_l. \tag{3}$$

An alternative Sen–Yates–Grundy formula for fixed sample size designs is (Sen, 1953; Yates, 1953):

$$\begin{aligned}\hat{V}\big(\hat{t}_{d\mathrm{HT}}\big) &= -\sum_{k\in s_d}\sum_{l<k;l\in s_d} a_{kl}(\pi_{kl}-\pi_k\pi_l)(a_k y_k - a_l y_l)^2 \\ &= \sum_{k\in s_d}\sum_{l<k;l\in s_d} (a_{kl}/a_k a_l - 1)(a_k y_k - a_l y_l)^2. \end{aligned}$$

These variance estimators are impractical because they contain second-order inclusion probabilities π_{kl} whose computation is often laborious for practical purposes. Hájek (1964) and Berger (2004, 2005b) proposed approximations to π_{kl}. Särndal (1996) developed efficient strategies with simple variance estimators under fixed sample size probability proportional-to-size (πPS) schemes, including a combination of Poisson sampling or stratified simple random sampling without replacement (SRSWOR) with

GREG estimation. Berger and Skinner (2005) proposed a jackknife variance estimator and Kott (2006a) introduced a delete-a-group jackknife variance estimator for πPS designs. The SAS procedure SURVEYSELECT is able to compute π_{kl} under certain unequal probability without-replacement sampling designs. Some software products can incorporate the π_{kl} into variance estimation procedures; an example is the SUDAAN software. The SAS macro CLAN includes the Sen–Yates–Grundy formula. Such estimators are discussed in Chapter 2.

Many πPS designs allow using of Hájek approximation (Berger, 2004, 2005b; Hájek, 1964) of second-order inclusion probabilities by $\pi_{kl} \approx \pi_k \pi_l \left[1 - (1 - \pi_k)(1 - \pi_l) m_d^{-1}\right]$ for $k \neq l$, where $m_d = \sum_{i \in U_d} \pi_i (1 - \pi_i)$. The approximation is used in a simple variance estimator $\hat{V}\left(\hat{t}_{d\text{HT}}\right) = \sum_{k \in s_d} c_k e_k^2$, where $c_i = n_d (n_d - 1)^{-1} (1 - \pi_i)$ and $e_k = a_k y_k - \left(\sum_{i \in s_d} c_i\right)^{-1} \sum_{i \in s_d} c_i a_i y_i$.

For unequal probability sampling designs, the variance of the ordinary HT estimator has been approximated under a with-replacement (WR) assumption, leading to Hansen–Hurwitz (1943) type variance estimator (Lehtonen and Pahkinen, 2004, p. 228, and SAS procedure SURVEYMEANS) given by

$$\hat{V}_A(\hat{t}_{d\text{HT}}) = \frac{1}{n_d(n_d - 1)} \sum_{k \in s_d} (n_d a_k y_k - \hat{t}_{d\text{HT}})^2. \tag{4}$$

For unplanned domains, the variance estimator for HT should account for random domain sizes. An approximate variance estimator applied, for example, in SAS procedure SURVEYMEANS contains extended domain variables y_{dk}:

$$\hat{V}_U(\hat{t}_{d\text{HT}}) = \frac{n}{n - 1} \sum_{k \in s} (a_k y_{dk} - \hat{t}_d / n)^2, \tag{5}$$

where n is the total sample size. Under SRSWOR, an alternative to (5) is

$$\hat{V}_{\text{srswor}}(\hat{t}_{d\text{HT}}) = N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) p_d \hat{s}_{dy}^2 \left(1 + \frac{q_d}{\text{c.v}_{dy}^2}\right),$$

where $p_d = n_{s_d}/n$, $q_d = 1 - p_d$, variance estimator is, $\hat{s}_{dy}^2 = \sum_{k \in s_d} (y_k - \bar{y}_d)^2 / (n_{s_d}$ $-1)$, and estimated coefficient of variation is $\text{c.v}_{dy} = \hat{s}_{dy} / \bar{y}_d$ for $\bar{y}_d = \sum_{k \in s_d} y_k / n_{s_d}$.

The HT estimator can be regarded as a model-dependent estimator under a model $Y_k = \beta \pi_k + \pi_k \varepsilon_k$ (Zheng and Little, 2003). HT is nearly optimal estimator among weighted sums of Y values when Y depends on scalar x as $E(Y_k) = \beta x_k$, the variance of errors is proportional to x_k^2, and the sampling design assigns π_k proportional to x_k. On the other hand, HT is very inefficient when the intercept of the model is far from zero. Disastrous results are possible in HT estimation, as the famous example of Basu (1971) shows (e.g., citation in Little, 2004).

If the domain size N_d is known, we expect better results with a "Hájek" type direct estimator $\hat{t}_{dH(N)} = N_d \hat{\bar{y}}_d$ (e.g., Hidiroglou and Patak, 2004; Särndal et al., 1992, p. 391) derived from the domain mean $\hat{\bar{y}}_d = \sum_{k \in s_d} a_k y_k / \hat{N}_d$ with $\hat{N}_d = \sum_{k \in s_d} a_k$. This is a special case of ratio estimation (Section 4.3.1). The variance of $\hat{t}_{dH(N)}$ is estimated by

$$\hat{V}(\hat{t}_{dH(N)}) = \left(\frac{N_d}{\hat{N}_d}\right)^2 \sum_{k \in s_d} \sum_{l \in s_d} (a_k a_l - a_{kl})(y_k - \hat{\bar{y}}_d)(y_l - \hat{\bar{y}}_d). \tag{6}$$

3.2. Population fit regression estimator

The population fit regression estimator is a theoretical tool used in approximating real-world estimators. We first consider *difference estimators* (Särndal, 1980; Särndal et al., 1992, p. 221). If known values y_k^0 are close to y_k, we write the estimable population total as

$$t = \sum_{k\in U} y_k = \sum_{k\in U} y_k^0 + \sum_{k\in U} (y_k - y_k^0).$$

A difference estimator is defined by estimating the second sum using HT:

$$\hat{t}_{\text{DIFF}} = \sum_{k\in U} y_k^0 + \sum_{k\in s} a_k(y_k - y_k^0).$$

As the y_k^0 are constants, $\hat{t}_{\text{DIFF}}$ is unbiased for t.

Consider a regression superpopulation model $Y_k = \mathbf{x}'_k\boldsymbol{\beta} + \varepsilon_k$, where $\mathbf{x}_k = (1, x_{1k}, \ldots, x_{Jk})'$ is the vector of auxiliary x-variables, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_J)'$ is the vector of regression coefficients, and ε_k are the residuals with variances $\sigma_k^2 = \text{Var}(\varepsilon_k)$. Hypothetically, we can fit the model to the population by calculating generalized least squares (GLS) estimator $\mathbf{B} = \hat{\boldsymbol{\beta}}$ as

$$\mathbf{B} = \left(\sum_{k\in U} \frac{\mathbf{x}_k\mathbf{x}'_k}{\sigma_k^2}\right)^{-1} \left(\sum_{k\in U} \frac{\mathbf{x}_k y_k}{\sigma_k^2}\right).$$

In practice, the error variance $\text{Var}(\varepsilon_k) = \sigma_k^2$ can often be assumed constant, $\sigma_k^2 = \sigma^2$, and then it cancels out. When the variance varies between observations, the σ_k^2 should be included in the estimators. Straightforward cases are known σ_k^2 or an assumption that the variances differ by known constants c_k such that $\sigma_k^2 = c_k\sigma^2$. A special case is when $c_k = 1$ for all $k \in U$. For more details on the treatment of σ_k^2, see, for example, Särndal et al. (1992, p. 229 and Chapter 7).

A difference estimator with fitted values $\hat{y}_k^0 = \mathbf{x}'_k\mathbf{B}$ defines the *population fit regression estimator*,

$$\hat{t}_{\text{REG}} = \sum_{k\in U} \hat{y}_k^0 + \sum_{k\in s} a_k(y_k - \hat{y}_k^0).$$

If an estimator $\hat{t}$ can be well approximated by $\hat{t}_{\text{REG}}$, then $\text{Var}(\hat{t})$ can be estimated by a sample-based estimator of

$$\text{Var}(\hat{t}_{\text{REG}}) = \text{Var}\left(\sum_{k\in s} a_k E_k\right) = \sum_{k\in U}\sum_{l\in U} (a_k a_l / a_{kl} - 1) E_k E_l,$$

where $E_k = y_k - \hat{y}_k^0$ are the population fit residuals. To estimate $\text{Var}(\hat{t}_{\text{REG}})$ from sample, we replace the E_k by corresponding sample residuals $e_k = y_k - \mathbf{x}'_k\hat{\mathbf{B}}$. If $\hat{\mathbf{B}}$ is nearly unbiased for $\mathbf{B}$, we can verify using $E(a_{kl}I_kI_l) = 1$ that a nearly unbiased estimator for $\text{Var}(\hat{t}_{\text{REG}})$ is

$$\hat{V}(\hat{t}_{\text{REG}}) = \sum_{k\in s}\sum_{l\in s} (a_k a_l - a_{kl}) e_k e_l. \tag{7}$$

One approach to estimate $\mathbf{B}$ is to plug in HT estimators of both of its sum components. When σ_k^2 is constant, we use a weighted least squares (WLS) estimator

$$\hat{\mathbf{B}} = \left(\sum_{k \in s} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \left(\sum_{k \in s} a_k \mathbf{x}_k y_k \right).$$

This estimator is only approximately unbiased due to its nonlinearity. Another approach is to consider the population maximum likelihood (ML) estimator maximizing $f(\boldsymbol{\beta}) = -\sum_{k \in U} \left(y_k - \mathbf{x}'_k \boldsymbol{\beta} \right)^2 / \sigma^2$. As only the sample is available, we use an estimated log-likelihood, the so-called *pseudolikelihood*, instead (Binder, 1983; Godambe and Thompson, 1986a; Nordberg, 1989). The function $f(\boldsymbol{\beta})$ is estimated by an unbiased HT type estimator $\hat{f}(\boldsymbol{\beta}) = -\sum_{k \in s} a_k \left(y_k - \mathbf{x}'_k \boldsymbol{\beta} \right)^2 / \sigma^2$. This function is maximized by $\hat{\mathbf{B}}$. Robust alternatives are presented in Beaumont and Alavi (2004).

Särndal et al. (1992) and Estevao and Särndal (2006) have approximated GREG and calibration estimators (Sections 3.3 and 3.4) by Taylor linearization yielding a population fit regression estimator. Because many approximations are involved, the resulting variance estimators are at least slightly biased.

3.3. *GREG estimators*

The GREG estimator is a sample-based substitute for the population fit regression estimator (Section 3.2). A direct type GREG estimator of domain total t_d is assisted by a regression model $Y_k = \mathbf{x}'_k \boldsymbol{\beta}_d + \varepsilon_k$, $\mathrm{Var}(\varepsilon_k) = \sigma_k^2$. Assuming constant error variance σ_k^2, the domain-specific parameter $\mathbf{B}_d$ of the population fit defined for U_d is estimated as in Section 3.2 by

$$\hat{\mathbf{B}}_d = \left(\sum_{k \in s_d} a_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \left(\sum_{k \in s_d} a_k \mathbf{x}_k y_k \right),$$

and the fitted values $\hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}_d$ and residuals $e_k = y_k - \hat{y}_k$ are incorporated into the GREG estimator

$$\hat{t}_{d\mathrm{GREG}} = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in s_d} a_k (y_k - \hat{y}_k) = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in s_d} a_k e_k \qquad (8)$$

(Särndal, 1980; Särndal et al., 1992). The first part in $\hat{t}_{d\mathrm{GREG}}$, the population sum of fitted values over the domain, is sometimes called a synthetic estimator (Särndal, 1984). When compared with direct GREG, it may have smaller variance but possibly large design bias. The weighted sum of residuals tends to correct for the design bias. In some cases, however, the weighted sum of the residual terms is zero. This happens when the model contains an intercept.

Rearranging the terms of GREG we obtain the traditional regression estimator

$$\hat{t}_{d\mathrm{GREG}} = \hat{t}_{d\mathrm{HT}} + (\mathbf{t}_{dx} - \hat{\mathbf{t}}_{dx})' \hat{\mathbf{B}}_d,$$

where $\mathbf{t}_{dx} = \sum_{k \in U_d} \mathbf{x}_k = \left(N_d, \sum_{k \in U_d} \mathbf{x}_{1k}, \ldots, \sum_{k \in U_d} \mathbf{x}_{Jk} \right)'$ and $\hat{\mathbf{t}}_{dx} = \sum_{k \in s_d} a_k \mathbf{x}_k$. By Taylor linearization, $\hat{t}_{d\mathrm{GREG}}$ is approximated by a population fit regression estimator

$\hat{t}_{d\text{REG}} = \hat{t}_{d\text{HT}} + (\mathbf{t}_{dx} - \hat{\mathbf{t}}_{dx})'\mathbf{B}_d$ applied in U_d. The estimator $\hat{t}_{d\text{REG}}$ is unbiased for t_d, and so the GREG estimator is nearly unbiased. Although GREG incorporates a model, it is model-assisted, not model-dependent, because the model only yields a fixed population quantity $\mathbf{B}_d$, and GREG is nearly design unbiased even when the model is not valid. By (7), the variance of $\hat{t}_{d\text{GREG}}$ can be estimated using sample residuals $e_k = y_k - \mathbf{x}'_k\hat{\mathbf{B}}_d$:

$$\hat{V}_1(\hat{t}_{d\text{GREG}}) = \sum_{k\in s_d}\sum_{l\in s_d}(a_k a_l - a_{kl})e_k e_l. \tag{9}$$

The GREG estimator can be written as a weighted sum of observations incorporating so-called g-weights:

$$\hat{t}_{d\text{GREG}} = \sum_{k\in s_d} a_k g_{dk} y_k;\ g_{dk} = I_{dk} + I_{dk}(\mathbf{t}_{dx} - \hat{\mathbf{t}}_{dx})'\hat{\mathbf{M}}_d^{-1}\mathbf{x}_k,$$

where $\hat{\mathbf{M}}_d = \sum_{i\in s_d} a_i\mathbf{x}_i\mathbf{x}'_i$ and $I_{dk} = I\{k \in U_d\}$ is the domain membership indicator. The g-weights are used in a variance estimator

$$\hat{V}_2(\hat{t}_{d\text{GREG}}) = \sum_{k\in s_d}\sum_{l\in s_d}(a_k a_l - a_{kl})g_{dk}e_k g_{dl}e_l \tag{10}$$

(Hidiroglou and Patak, 2004; Särndal et al., 1989 and 1992, p. 235). In practice, $\hat{V}_1$ and $\hat{V}_2$ often yield similar results but $\hat{V}_2$ in (10) is preferable (Fuller, 2002; Särndal et al., 1989).

3.4. Calibration estimators

Calibration is based on information about known totals of auxiliary variables $\mathbf{x}_k$, also called *benchmark variables*, at an aggregate level. In model-free calibration (Särndal, 2007) discussed here, it is not necessary to impose a model on the data. Suppose the population is divided into *calibration groups* U_c $(c = 1, 2, \ldots, C)$ so that every domain U_d is contained within one of the groups and the population totals $\mathbf{t}_{cx} = \sum_{k\in U_c}\mathbf{x}_k$ of auxiliary variables are known. The domain totals $\mathbf{t}_{dx}$ are not required. Direct *calibration estimator* of the domain total t_d is a weighted sum of observations:

$$\hat{t}_{d\text{CAL}} = \sum_{k\in s_d} w_k y_k,$$

where the *calibration weights* w_k have to satisfy the *calibration equation*s

$$\sum_{k\in s_c} w_k\mathbf{x}_k = \sum_{k\in U_c}\mathbf{x}_k = \mathbf{t}_{cx}$$

for every calibration group. It follows immediately that calibration estimator applied to the auxiliary data yields the known totals. We therefore expect that the weighted sum of y over s_d is close to t_d.

There are two main approaches to calibration, one based on a *distance measure* and the other based on *instrument vectors* (Chapter 25). In the distance measure approach, the weights w_k minimize a distance to the design weights a_k, subject to the calibration equations (Deville and Särndal, 1992; Singh and Mohl, 1996). An example of a calibration estimator incorporating an instrument vector $\mathbf{z}_k$ is

$$\hat{t}_{d\text{CAL}} = \sum_{k \in s_d} a_k (1 + \boldsymbol{\lambda}' \mathbf{z}_k) y_k,$$

where $\boldsymbol{\lambda}' = (\mathbf{t}_{cx} - \hat{\mathbf{t}}_{cx})' \left(\sum_{k \in s_c} a_k \mathbf{z}_k \mathbf{x}'_k \right)^{-1}$. It should be noted that the values of instrument z-variables need to be known only for the sample (or need to be estimated); they are not necessarily treated as proper auxiliary information in the same manner as the auxiliary x-variables. For practical purposes, a natural choice is $\mathbf{z}_k = \mathbf{x}_k$; an optimal choice is discussed in Estevao and Särndal (2004).

As in (7), the variance of $\hat{t}_{d\text{CAL}}$ is estimated by

$$\hat{V}(\hat{t}_{d\text{CAL}}) = \sum_{k \in s_c} \sum_{l \in s_c} (a_k a_l - a_{kl})(y_{dk} - \mathbf{x}'_{ck} \hat{\mathbf{B}}_{cd})(y_{dl} - \mathbf{x}'_{cl} \hat{\mathbf{B}}_{cd}),$$

where $\mathbf{x}_{ck} = I\{k \in U_c\} \mathbf{x}_k$ (Estevao and Särndal, 2006), and

$$\hat{\mathbf{B}}_{cd} = \left(\sum_{k \in s_c} a_k \mathbf{z}_k \mathbf{x}'_{ck} \right)^{-1} \left(\sum_{k \in s_c} a_k \mathbf{z}_k y_{dk} \right).$$

When U_c is much larger than U_d, the variance can become large. Therefore, we should attempt to find a calibration group that agrees closely with the domain of interest.

Our GREG estimator of Section 3.3 is actually a special case of calibration, sometimes called linear calibration estimator, as the weights $a_k g_{dk}$ minimize a certain chi-square distance to design weights a_k, subject to domain-level calibration equations $\sum_{k \in s_d} a_k g_{dk} \mathbf{x}_k = \mathbf{t}_{dx}$.

Calibration is contrasted with GREG estimation in Särndal (2007). Särndal and Lundström (2005) discuss calibration in the context of adjustment for unit nonresponse in sample surveys.

3.5. *Computational example with direct estimation under a planned domain structure*

In this section, we demonstrate with real data the direct Horvitz–Thompson, Hájek, and GREG estimation of totals for domains. The data set contains disposable income of households in $D = 12$ regions of Western Finland. The population consists of $N = 431{,}000$ households. In addition to the income data, the record of a household shows the number of household members who had higher education (variable EDUC) and the number of months in total the household members were employed (EMP) during last year. All three variables were determined using administrative registers. For this computational exercise, we had access to population level information on all variables. This gives a possibility to compare sample estimates to the known population values.

We were interested in the yearly total disposable income $t_d = \sum_{k \in U_d} y_k$ in the regions $U_d (d = 1, \ldots, D)$. A sample of 1000 households was drawn from the population by using stratified πPS (without-replacement type probability proportional to size sampling) with household size as the size variable. To demonstrate estimation for planned domains, we interpret here the sample as a stratified sample where the regions constitute the strata. Thus, the domain structure is of planned type, where the regional sample sizes are considered fixed by the sampling design. In Section 4.2, we use the same sample in estimation for unplanned domains, where the regional sample sizes are considered random.

In Table 2, we grouped the domains by sample size into minor ($8 \leq n_d \leq 33$), medium-sized ($34 \leq n_d \leq 45$) and major ($46 \leq n_d \leq 277$) domains, where n_d is the observed domain sample size in domain U_d. There were four domains in each domain size class.

Results are shown in Table 2. The absolute relative error of an estimator in domain d is calculated as $|\hat{t}_d - t_d|/t_d$ and domain group's MARE is the mean of absolute relative errors over domains in the group. Correspondingly, MCV is the mean coefficient of variation of the estimate over domain group. The coefficient of variation is calculated as $\text{s.e}(\hat{t}_d)/\hat{t}_d$, where s.e refers to the estimated standard error of an estimator. For variance estimation, we approximated the design by with-replacement type probability-proportional-to-size sampling (PPS). The variance estimators for ordinary HT (column 1) and the Hájek type estimator (column 2) were defined by (4) and (6), respectively. The Hájek estimator, which contains the known domain sizes N_d, yielded better results than ordinary HT.

A calibration estimator, the direct GREG estimator with linear assisting model,

$$Y_k = \beta_{0d} + \beta_{1d}\text{EMP}_k + \varepsilon_k(\text{column 3}) \text{ or}$$
$$Y_k = \beta_{0d} + \beta_{1d}\text{EMP}_k + \beta_{2d}\text{EDUC}_k + \varepsilon_k(\text{column 4}),$$

and variance estimator (10) incorporated the known domain sizes and domain totals of EMP (column 3) and EDUC (column 4). The model parameters were estimated by WLS with weights $a_k = 1/\pi_k$. By GREG, we obtained clearly smaller MARE and MCV figures than by HT.

Adding information in the estimation procedure improved the results until the assisting model contained both EMP and EDUC: inclusion of EDUC in GREG decreased MCV but average errors did not always decrease. In large domains, the average error and MCV were usually smaller than in small domains.

Table 2
Mean absolute relative error (MARE) and mean coefficient of variation (MCV) of direct HT, Hájek, and calibration (GREG) estimators of totals for minor, medium-sized, and major domains by using various amounts of auxiliary information in a planned domains case

	HT		Hájek		Calibration (GREG)			
Auxiliary Information	1 None		2 Domain Sizes		3 Domain Sizes and Domain Totals of EMP		4 Domain Sizes and Domain Totals of EMP and EDUC	
Domain sample size class	MARE (%)	MCV (%)	MARE (%)	MCV (%)	MARE (%)	MCV (%)	MARE (%)	MCV (%)
Minor $8 \leq n_d \leq 33$	11.5	11.9	5.3	10.9	5.8	7.7	6.4	6.8
Medium $34 \leq n_d \leq 45$	7.6	9.0	6.4	9.0	3.7	8.0	3.6	8.1
Major $46 \leq n_d \leq 277$	12.5	5.2	4.7	5.6	4.3	4.7	5.2	3.7