Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Use of within-primary-sample-unit variances to assess the stability of a standard design-based variance estimator

by Donsig Jang and John L. Eltinge

December 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2009
Vol. 35, No. 2, pp. 235-245
Statistics Canada, Catalogue No. 12-001-X

Use of within-primary-sample-unit variances to assess the stability of a standard design-based variance estimator

Donsig Jang and John L. Eltinge [1]

Abstract

In analysis of sample survey data, degrees-of-freedom quantities are often used to assess the stability of design-based variance estimators. For example, these degrees-of-freedom values are used in construction of confidence intervals based on t distribution approximations; and of related t tests. In addition, a small degrees-of-freedom term provides a qualitative indication of the possible limitations of a given variance estimator in a specific application. Degrees-of-freedom calculations sometimes are based on forms of the Satterthwaite approximation. These Satterthwaite-based calculations depend primarily on the relative magnitudes of stratum-level variances. However, for designs involving a small number of primary units selected per stratum, standard stratum-level variance estimators provide limited information on the true stratum variances. For such cases, customary Satterthwaite-based calculations can be problematic, especially in analyses for subpopulations that are concentrated in a relatively small number of strata. To address this problem, this paper uses estimated within-primary-sample-unit (within PSU) variances to provide auxiliary information regarding the relative magnitudes of the overall stratum-level variances. Analytic results indicate that the resulting degrees-of-freedom estimator will be better than modified Satterthwaite-type estimators provided: (a) the overall stratum-level variances are approximately proportional to the corresponding within-stratum variances; and (b) the variances of the within-PSU variance estimators are relatively small. In addition, this paper develops errors-in-variables methods that can be used to check conditions (a) and (b) empirically. For these model checks, we develop simulation-based reference distributions, which differ substantially from reference distributions based on customary large-sample normal approximations. The proposed methods are applied to four variables from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Key Words: Complex sample design; Degrees of freedom; Errors-in-variables regression; Satterthwaite approximation; Stratified multistage sample survey; Two-PSU-per-stratum design; U.S. Third National Health and Nutritional Examination Survey (NHANES III).

1. Introduction

1.1 Motivating example: Inference for special subpopulations in NHANES III

This work arose from a study of inference for geographically concentrated subpopulations in the U.S. Third National Health and Nutrition Examination Survey (NHANES III). For some general background on NHANES III, see National Center for Health Statistics (1996). In many analyses, NHANES III data are treated as arising from a stratified multistage sample design that uses 49 strata and two primary sample units (PSUs) per stratum. Consequently, formal inferences from NHANES III data (*e.g.*, construction of confidence intervals) often use the assumption that the associated variance estimators are based on approximately 49 degrees of freedom and are thus relatively stable.

However, the Mexican-American subpopulation is concentrated in a relatively small number of strata, so associated variance estimators may be less stable (*i.e.*, have greater sampling variability) than would be indicated by the nominal 49 degrees of freedom term. Consequently, it is important to use an appropriate estimator of the true degrees of freedom associated with variance estimators for such subpopulations, and to modify confidence interval calculations accordingly. Development of an appropriate degrees-of-freedom estimator can be complicated by moderate or severe heterogeneity in the underlying stratum-level variances. Such complications arose in the analysis of the four NHANES III variables listed in Table 1.1. Section 5 will consider inference for the means of these four variables for the subpopulation of Mexican-Americans aged 20-29.

Table 1.1
Four NHANES III variables

Variable Name	Description
BMPWT	Weight (*kg*)
HAR3	Do you smoke cigarettes now? (0/1)
TCRESULT	Serum total cholesterol (*mg/dL*)
HDRESULT	HDL cholesterol (*mg/dL*)

1.2 Stability of design-based variance estimators

Suppose we have a population partitioned into L strata, with N_h PSUs in stratum h for $h = 1, 2, ..., L$. Under a

1. Donsig Jang, Mathematica Policy Research, 600 Maryland Avenue SW, Suite 550, Washington, DC 20024-2512, U.S.A. E-mail: DJang@Mathematica-mpr.com; John L Eltinge, U.S. Bureau of Labor Statistics, PSB 1950, 2 Massachusetts Avenue NE, Washington, DC 20212-0001, U.S.A. E-mail: Eltinge_J@bls.gov.

stratified multistage sampling design, we select n_h PSUs, with replacement, and with per-draw selection probability p_{hi} for PSU i within stratum h where $\sum_{i=1}^{N_h} p_{hi} = 1$. Thus, a total of $n = \sum_{h=1}^{L} n_h$ PSUs are selected. Within selected PSU (h, i), n_{hi} secondary sample units (SSUs) are selected with replacement and with per-draw selection probabilities p_{hij}, where $\sum_{j=1}^{N_{hi}} p_{hij} = 1$ and N_{hi} is the number of SSUs in PSU (h, i). For a given survey item, let Y_h be the population total for stratum h, and define the overall population total $Y = \sum_{h=1}^{L} Y_h$. The total Y may correspond to a total either for the full population or for a specified subpopulation.

Our goal is to construct a confidence interval for the total Y. Let $\hat{Y}_{hij}$ be an unbiased estimator of Y_{hij}, the population total for secondary unit j in primary unit i in stratum h. Then a customary design-based estimator of Y is $\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h$, where $\hat{Y}_h = n_h^{-1} \sum_{i=1}^{n_h} p_{hi}^{-1} \hat{Y}_{hi}$; $p_{hi}^{-1} \hat{Y}_{hi}$ is a design unbiased estimator of Y_h based on data obtained from PSU i in stratum h; and $\hat{Y}_{hi} = n_{hi}^{-1} \sum_{j=1}^{n_{hi}} p_{hij}^{-1} \hat{Y}_{hij}$ is an unbiased estimator of Y_{hi}, the population total for PSU i in stratum h.

Under the standard condition that sampling is independent across strata, the variance of $\hat{Y}$ can be written, $V(\hat{Y}) = \sum_{h=1}^{L} V_h$ where $V_h = \mathrm{Var}(\hat{Y}_h)$. Throughout the remainder of this paper, we will call the V_h terms the stratum-level variances, and we will assume that $n_h \geq 2$ for all $h = 1, 2, ..., L$. Note that V_h depends on the sample design used within stratum h, and is distinct from the within-stratum variance of element-level Y values. A simple unbiased estimator for $V(\hat{Y})$ is $\hat{V}(\hat{Y}) = \sum_{h=1}^{L} \hat{V}_h$ where $\hat{V}_h = n_h^{-1} (n_h - 1)^{-1} \sum_{h=1}^{n_h} (p_{hi}^{-1} \hat{Y}_{hi} - \hat{Y}_h)^2$; see, *e.g.*, Wolter (1985, page 44). Note that the estimator $\hat{V}_h$ is a multiple of a sum of squared differences among the terms $p_{hi}^{-1} \hat{Y}_{hi}$. In addition, under regularity conditions the random variables $p_{hi}^{-1} \hat{Y}_{hi}$ will be approximately normally distributed for a given stratum h. Consequently, the overall stratum-level variance estimators $\hat{V}_h$ generally will approximately satisfy the following condition.

(C.1) For $h = 1, 2, ..., L$, the terms $V_h^{-1} (n_h - 1) \hat{V}_h$ are distributed as independent chi-square random variables with $n_h - 1$ degrees of freedom, respectively, where $n_h \geq 2$.

Under condition (C.1), $\{V(\hat{Y})\}^{-1} d \hat{V}(\hat{Y})$ has the same first and second moments as a chi-square random variable with d degrees of freedom, where d is the solution to the equation,

$$2\{V(\hat{Y})\}^2 - V\{\hat{V}(\hat{Y})\}\, d = 0 \tag{1.1}$$

or equivalently

$$d \overset{\text{def}}{=} \left\{ \sum_{h=1}^{L} (n_h - 1)^{-1} V_h^2 \right\}^{-1} \{V(\hat{Y})\}^2 \tag{1.2}$$

where $V\{\hat{V}(\hat{Y})\} = \sum_{h=1}^{L} 2 (n_h - 1)^{-1} V_h^2$. Direct substitution of $\hat{V}_h$ for V_h and $\hat{V}(\hat{Y})$ for $V(\hat{Y})$ in expression (1.2) leads to the Satterthwaite (1946)-type degrees-of-freedom estimator,

$$\hat{d}_S = \left\{ \sum_{h=1}^{L} (n_h - 1)^{-1} \hat{V}_h^2 \right\}^{-1} \{\hat{V}(\hat{Y})\}^2. \tag{1.3}$$

For some general background on $\hat{d}_S$ and related estimators, see, *e.g.*, Smith (1936), Satterthwaite (1941, 1946), Cochran (1977, page 96) and Kendall, Stuart and Ord (1983, pages 91-92). In constructing confidence intervals for a subpopulation parameter, Casady, Dorfman and Wang (1998) use Bayesian ideas to develop related degrees-of-freedom measures for a Student's t-statistic.

For designs in which n_h is large for all h, the error in estimation of V_h is relatively small, and $\hat{d}_S$ can provide a satisfactory estimator of expression (1.2). However, many large-scale surveys use small n_h, *e.g.*, $n_h = 2$. For small-n_h cases, condition (C.1) and routine algebra lead to the expectation result $E(\hat{V}_h^2) = (n_h - 1)^{-1} (n_h + 1) V_h^2$. This implies that the standard Satterthwaite degrees-of-freedom estimator $\hat{d}_S$ can severely underestimate d, and that the corresponding confidence interval $\hat{Y} \pm t_{\hat{d}_S, 1-\alpha/2} \{\hat{V}(\hat{Y})\}^{1/2}$ may have a true coverage rate substantially below the nominal rate $1 - \alpha$. Consequently, Jang (1996) considered an alternative degrees-of-freedom estimator,

$$\hat{d}_{mS} = (3L + 14)^{-1} (9L)\, \hat{d}_S. \tag{1.4}$$

for the two-PSUs-per-stratum design.

1.3 Use of auxiliary stratum-level data

For cases in which there is moderate heterogeneity among the V_h terms, simulation work by Jang (1996) indicated that $\hat{d}_{mS}$ performs relatively well. However, if there is substantial heterogeneity among the stratum variances (*i.e.*, if $L^{-1} d$ is relatively small), then $\hat{d}_{mS}$ may be unsatisfactory. The fundamental problem is that when the n_h values are relatively small, the estimators $\hat{V}_h$, by themselves, do not provide sufficient information regarding the relative magnitudes of the true stratum-level variances V_h. In some cases, a variance estimator based on auxiliary data is expected to be more stable than the customary design-based estimator; see *e.g.*, Isaki (1983). Similarly, auxiliary sources of information can be used to evaluate the relative magnitudes of the variances V_h.

The remainder of this paper will focus on auxiliary information provided by relationships between the overall stratum-level variances V_h and associated within-PSU variances. Recall from Wolter (1985, page 41) the decomposition,

$$\mathrm{Var}(\hat{Y}_h) = V_{Bh} + V_{Wh}, \tag{1.5}$$

where $V_{Bh} = \mathrm{Var}\{\sum_{i=1}^{n_h}(n_h p_{hi})^{-1} Y_{hi}\}$ is the between-PSU variance, $V_{Wh} = \sum_{i=1}^{N_h}(n_h p_{hi})^{-1}\sigma_{2hi}^2$ is the within-PSU variance, $Y_{hi} = E(\hat{Y}_{hi} \mid \text{PSU } i, \text{ stratum } h)$ and $\sigma_{2hi}^2 = \mathrm{Var}(\hat{Y}_{hi} \mid \text{PSU } i,$ stratum $h)$. In addition, define $\overline{V}_W = L^{-1}\sum_{h=1}^{L} V_{Wh}$.

Estimators of V_{Wh} can provide useful auxiliary information on the relative magnitudes of V_h for two reasons. First, for designs with a small n_h and relatively large n_{hi}, the within-PSU variance estimators $\hat{V}_{Wh}$ may be considerably more stable than $\hat{V}_h$. Second, in some applications (*e.g.*, some of the examples presented in Section 5 below), observed variance estimates are consistent with a model under which V_h is proportional to V_{Wh}, *i.e.*,

$$V_h = \beta_1 V_{Wh} \text{ for all } h = 1, \ldots, L, \tag{1.6}$$

where β_1 is a fixed constant. The proportionality relationship (1.6) would arise if both V_{Bh} and V_{Wh} are proportional to a common scale factor, *e.g.*, $(\overline{Y}_h)^{\alpha}$ for some power α. Under relationship (1.6), expression (1.2) may be rewritten,

$$d = \left\{\sum_{h=1}^{L}(n_h - 1)^{-1} V_{Wh}^2\right\}^{-1}\left\{\sum_{h=1}^{L} V_{Wh}\right\}^2. \tag{1.7}$$

Consequently, given a set of stable within-PSU variance estimators $\hat{V}_{Wh}$ and associated variance-of-variance-estimators $\widehat{\mathrm{Var}}(\hat{V}_{Wh})$,

$$\hat{d}_{WS} = \left\{\sum_{h=1}^{L}(n_h - 1)^{-1}[\hat{V}_{Wh}^2 - \widehat{\mathrm{Var}}(\hat{V}_{Wh})]\right\}^{-1}\left(\sum_{h=1}^{L}\hat{V}_{Wh}\right)^2 \tag{1.8}$$

is an alternative estimator of d.

Section 2 considers some of the properties of $\hat{d}_{WS}$. Section 3.1 uses errors-in-variables tests to check the adequacy of the proportionality condition (1.6). Section 3.2 presents two related diagnostics for the relationship between V_h and auxiliary variables, and for the magnitude of the error in the observed auxiliary variables $\hat{V}_{Wh}$.

A simulation study in Section 4 explores conditions under which the proposed new estimator $\hat{d}_{WS}$ may perform better than $\hat{d}_{mS}$. This assessment considers both the estimation of d as such, and the performance of confidence intervals for Y. Section 5 applies the proposed estimator to four variables from NHANES III, with emphasis on cases for which differences between the proposed estimators $\hat{d}_{WS}$ and $\hat{d}_{mS}$ have a substantial practical effect on assessment of the stability of the variance estimator $\hat{V}(\hat{Y})$. Section 6 reviews the methods developed in this paper and considers some possible extensions.

2. An estimator based on auxiliary information

2.1 A within-PSU variance estimator

A simple estimator of V_{Wh} is

$$\hat{V}_{Wh} = n_h^{-2}\sum_{i=1}^{n_h} p_{hi}^{-2}\,\hat{\sigma}_{2hi}^2, \tag{2.1}$$

where $\hat{\sigma}_{2hi} = n_{hi}^{-1}(n_{hi} - 1)^{-1}\sum_{j=1}^{n_{hi}}(p_{hij}^{-1}\hat{Y}_{hij} - \hat{Y}_{hi})^2$. Note that $\hat{\sigma}_{2hi}^2$ is approximately unbiased for σ_{2hi}^2 under a with-replacement sampling design within PSU i in stratum h; or under simple random sampling without replacement and with a small sampling fraction, $f_{hi} = N_{hi}^{-1} n_{hi}$. Standard sampling theory shows that $\hat{V}_{Wh}$ is approximately unbiased for V_{Wh}. Then an approximately unbiased estimator of $\mathrm{Var}(\hat{V}_{Wh})$ is

$$\widehat{\mathrm{Var}}(\hat{V}_{Wh}) = n_h^{-1}(n_h - 1)^{-1}\sum_{i=1}^{n_h}(\hat{V}_{Whi} - \hat{V}_{Wh})^2, \tag{2.2}$$

where $\hat{V}_{Whi} = n_h^{-1} p_{hi}^{-2}\hat{\sigma}_{2hi}^2$; see, *e.g.*, Eltinge and Jang (1996) and references cited therein. Note that the overall stratum-level variance estimators $\hat{V}_h$ are functions of the sample means of $p_{hij}^{-1}\hat{Y}_{hij}$ over PSUs in stratum h. In addition, the estimators $\hat{V}_{Wh}$ are functions of sample variances of the $p_{hij}^{-1}\hat{Y}_{hij}$ within the PSU (h, i). Thus, for variables Y for which $p_{hij}^{-1}\hat{Y}_{hij}$ are approximately normally distributed within stratum h, the estimators $\hat{V}_h$ and $\hat{V}_{Wh}$ are approximately independent.

2.2 Properties of $\hat{d}_{WS}$

In the remainder of this paper, the estimator $\hat{d}_{WS}$ defined in expression (1.8) will use $\widehat{\mathrm{Var}}(\hat{V}_{Wh})$ as defined in expression (2.2). Also, the remainder of this paper will use several asymptotic results. These results will use the condition that the number of strata, L, is increasing, while stratum-level PSU and SSU sample sizes n_h and m_h are allowed to remain small. This is in keeping with many practical multi-stage designs that use $n_h = 2$ and moderate values of m_h. See, *e.g.*, Krewski and Rao (1981) for a detailed development of large-L asymptotic results. The proof of Result 2.1 is routine and is thus omitted.

Result 2.1. Assume that $E(\hat{V}_{Wh}^r) = O(1)$ for $r = 1, 2, 3, 4$ and define

$$\hat{\overline{V}}_W = L^{-1}\sum_{h=1}^{L}\hat{V}_{Wh}$$

and (2.3)

$$\hat{\overline{V}}_{w(2)} = L^{-1}\sum_{h=1}^{L}(n_h - 1)^{-1}\{\hat{V}_{Wh}^2 - \widehat{\mathrm{Var}}(\hat{V}_{Wh})\}.$$

Then $\hat{\hat{V}}_W$ and $\hat{\hat{V}}_{W(2)}$ are consistent estimators of $\bar{V}_W$ and $L^{-1}\sum_{h=1}^{L}(n_h-1)^{-1}V_{Wh}^2$, respectively. In addition, $L^{-1}\hat{d}_{WS}$ is a consistent estimator of $L^{-1}d_{WS}$.

Section 1 suggested that in some cases, the auxiliary-data based estimator $\hat{d}_{WS}$ might be more stable than the modified Satterthwaite estimator $\hat{d}_{mS}$. To examine this idea, we will compare the variances of $\hat{d}_{WS}$ and $\hat{d}_{mS}$ under condition (C.1) and the following additional assumptions.

(C.2) For $h = 1, 2, ..., L$, $V_{Wh}^{-1}(m_h-1)\hat{V}_{Wh}$ are distributed as independent chi-square random variables with $m_h - 1$ degrees of freedom, respecttively, where m_h is the number of SSUs in stratum h; and are mutually independent of $\hat{V}_h$.

(C.3) For all $h = 1, 2, ..., L$, $n_h = 2$; and $m_h = m_0$ for some fixed positive integer $m_0 \geq 2$.

Arguments similar to those for condition (C.1) indicate that condition (C.2) may be satisfied approximately if within a given PSU (h, i), the m_h random variables $p_{hij}^{-1}\hat{Y}_{hij}$ are approximately independent and identically distributed normal random variables. Condition (C.3) restricts attention to the common case $n_h = 2$. In addition, condition (C.3) requires that an equal number, m_0, of secondary units be selected within each selected PSU. This allows simplification of the resulting approximations for the variances of $\hat{d}_{WS}$, as presented in Result 2.2.

Result 2.2. Assume conditions (C.1), (C.2), (C.3), and (1.6), and define $a = 4\mu_{A_2}^2\mu_{B_2}^{-2}\mathrm{Var}(A_2)$, $b = 4\mu_{A_2}^3\mu_{B_2}^{-3}\mathrm{Cov}(A_2, B_2)$, and $c=\mu_{A_2}^4\mu_{B_2}^{-4}\widehat{\mathrm{Var}(B_2)}$, where $A_2 = L^{-1}\sum_{h=1}^{L}\hat{V}_{Wh}$, $B_2 = L^{-1}\sum_{h=1}^{L}\{\hat{V}_{Wh}^2 - \widehat{\mathrm{Var}}(\hat{V}_{Wh})\}$, $\mu_{A2}=\bar{V}_W$ and $\mu_{B2}=L^{-1}\sum_{h=1}^{L}V_{Wh}^2$. Then

(i) the variances of the leading terms in Taylor expansions of $L^{-1}(\hat{d}_{WS}-d)$ and $L^{-1}(\hat{d}_{mS}-d)$ are, respectively,

$$\mathrm{V}_{LW} = a - b + c$$

and

$$\mathrm{V}_{Lm} = \frac{1}{9}\left(\frac{9L}{3L+14}\right)^2 (m_0-1)\left\{a-b+\frac{4(m_0-1)}{3(m_0+2)}c\right\}.$$

(ii) for all $m_0 \geq \lim_{L\to\infty} g(a,b,c)$, $\lim_{L\to\infty} V_{Lm} \geq \lim_{L\to\infty} V_{LW}$ where

$$g(a, b, c) = \{2(3a-3b+4c)\}^{-1}\left\{11c+\sqrt{144a^2+144b^2+153c^2-288ab+216ac-216bc}\right\}.$$

(iii) for $m_0 \geq 10$, $\lim_{L\to\infty} V_{Lm} \geq \lim_{L\to\infty} V_{LW}$ regardless of the values of the limiting moments $\lim_{L\to\infty}(\mu_{A_2}, \mu_{B_2}, L^{-1}\sum_{h=1}^{L}V_{Wh}^3, L^{-1}\sum_{h=1}^{L}V_{Wh}^4)$.

Result 2.2 indicates that for large L, $\hat{d}_{WS}$ may be preferable to $\hat{d}_{mS}$, provided: (1) the proportionality condition (1.6) is satisfied; and (2) the secondary unit sample size m_0 exceeds the lower bound given by $g(a,b,c)$ (thus ensuring relatively small variances of the $\hat{V}_{Wh}$). This motivates the use of within-PSU variances to assess the stability of survey variance estimators, especially under sample designs with small numbers of PSUs per stratum. For some additional discussion of this point, and some specific diagnostics to check the stability of $\hat{V}_{Wh}$, see Eltinge and Jang (1996) and references cited therein. For the four cases considered in Table 1.1 and studied further in Section 4 below, $g(a,b,c)$ is equal to 4.7, 4.3, 4.6, and 4.8 respectively, while the NHANES III application had the mean of the m_h values approximately equal to 22. In addition, we are treating V_{Wh} values as fixed, and Result 2.2 depends on the limiting moments of these V_{Wh} terms. Suppose that $V_{Wh}/\bar{V}_W$ had the same moments as F/f, where F follows a chi-square distribution on f degrees of freedom. Then $f = \infty$ corresponds to the case in which $V_{Wh} = \bar{V}_W$ for all h, which corresponds to the case in which the true d in (1.1) equals the customary value of $n - L$.

3. Testing the proportionality condition

3.1 An errors-in-variables model for V_h and V_{Wh}

Development of the alternative estimator $\hat{d}_{WS}$ in Section 1, and evaluation of its properties in Section 2, depended heavily on the proportionality condition (1.6). One may test the adequacy of this condition through the following steps. First, note that condition (1.6) is a special case of the following model,

(C.4) For all $h = 1, 2, ..., L$,

$$V_h = \beta_0 + \beta_1 V_{Wh} + q_h \tag{3.1}$$

where β_0 and β_1 are constants, and q_h is an equation error with mean zero and variance σ_{qqh}.

Second, recall that V_h and V_{Wh} are unknown quantities, for which we have the unbiased estimators $\hat{V}_h$ and $\hat{V}_{Wh}$, respectively. Using the errors-in-variables model notation in Fuller (1987), define the estimation errors

$$e_h = \hat{V}_h - V_h \qquad \text{and} \qquad u_h = \hat{V}_{Wh} - V_{Wh}. \tag{3.2}$$

Under conditions (C.1) and (C.2), the vector $(e_h, u_h)'$ is distributed with a mean vector equal to $(0,0)'$ and a variance-covariance matrix equal to $\text{diag}(\sigma_{eeh}, \sigma_{uuh})$ where $\sigma_{eeh} = (n_h - 1)^{-1} 2V_h^2$ and $\sigma_{uuh} = (m_h - 1)^{-1} 2V_{wh}^2$. Under the additional condition (C.3), these variance terms simplify to $\sigma_{eeh} = 2V_h^2$ and $\sigma_{uuh} = (m_0 - 1)^{-1} 2V_{Wh}^2$.

Expressions (3.1) and (3.2) define an errors-in-variables regression model with heterogeneous measurement error variances and non-normal errors. In addition $\widehat{\text{Var}}(\hat{V}_{Wh})$ defined in expression (2.2) is an unbiased estimator of σ_{uuh}, and thus provides identifying information for the parameters β_0, β_1 and σ_{qqh} in model (3.1) – (3.2). A direct application of Fuller (1987, pages 187-189) with equal weights then gives the consistent estimators (for increasing L),

$$\hat{\beta}_0 = L^{-1} \sum_{h=1}^{L} \hat{V}_h - \hat{\beta}_1 \hat{\bar{V}}_W,$$

$$\hat{\beta}_1 = \left[\sum_{h=1}^{L} (\hat{V}_{Wh} - \hat{\bar{V}}_W)^2 - \hat{\sigma}_{uu.} \right]^{-1} \sum_{h=1}^{L} (\hat{V}_{Wh} - \hat{\bar{V}}_W) \hat{V}_h, \quad (3.3)$$

and

$$\hat{\sigma}_{qq.} = \max \left[0,\ L^{-1} \sum_{h=1}^{L} (n_h - 1)^{-1} \{ (L-2)^{-1} L (\hat{V}_h - \hat{\beta}_0 - \hat{\beta}_1 \hat{V}_{Wh})^2 - (\hat{\sigma}_{eeh} + \hat{\beta}_1^2 \hat{\sigma}_{uuh}) \} \right], \quad (3.4)$$

where

$$\hat{\sigma}_{uu.} = \sum\nolimits_{h=1}^{L} \widehat{\text{Var}}(\hat{V}_{Wh}),\ \hat{\bar{V}}_W = L^{-1} \sum\nolimits_{h=1}^{L} \hat{V}_{Wh},$$

and (3.5)

$$\hat{\sigma}_{eeh} = 2(n_h + 1)^{-1} \hat{V}_h^2$$

from condition (C.1). In addition, direct application of Fuller (1987, page 188) leads to variance estimators $\hat{V}(\hat{\beta}_0)$ and $\hat{V}(\hat{\beta}_1)$, say; details are available from the authors.

3.2 Two related diagnostics

In keeping with condition (C.4), the proposed estimator $\hat{d}_{WS}$ is intended for cases in which the $\hat{V}_{Wh}$ provide useful auxiliary information on the relative magnitudes of the overall stratum-level variances V_h. To identify such cases, one simple diagnostic is the ratio $\{\hat{V}(\hat{V}_h)\}^{-1} \{\hat{\beta}_1^2 \hat{V}(\hat{V}_{Wh}) + \hat{\sigma}_{qqh}\}$, *i.e.*, the ratio of estimators of the variances of the approximate distributions of $\hat{V}_h - V_h$ and $\beta_1 \hat{V}_{Wh} - V_h$, respectively, under model (3.1) – (3.2). If this ratio is substantially less than unity, then use of $\hat{d}_{WS}$ may be indicated.

In addition, the performance of the estimator $\hat{d}_{WS}$ depends heavily on the magnitude of $\hat{\sigma}_{uu.}$ relative to the variability of the true within-PSU variances V_{Wh}. Define an estimator of the reliability ratio (Fuller 1987, page 3)

$$\hat{\kappa}_{xx} = \max \left\{ 0, \left[\sum_{h=1}^{L} (\hat{V}_{Wh} - \hat{\bar{V}}_W)^2 \right]^{-1} \left[\sum_{h=1}^{L} (\hat{V}_{Wh} - \hat{\bar{V}}_W)^2 - \hat{\sigma}_{uu.} \right] \right\}.$$

The values of $\hat{\kappa}_{xx}$ are between 0 and 1; and values of $\hat{\kappa}_{xx}$ close to unity indicate relatively small errors in the estimation of within-PSU variances. Conversely, small values of $\hat{\kappa}_{xx}$ (*e.g.*, $\hat{\kappa}_{xx} < 0.7$) may indicate that the methods of Sections 3.1 – 3.2 may not perform well, due to the relatively large sampling errors in the auxiliary information $\hat{V}_{Wh}$. The numerical work in Sections 4 and 5 below will consider these diagnostics further.

The work in this section is based on the assumption that $\sigma_{qq.} > 0$. One may develop related diagnostics applicable to the case of no equation errors, *i.e.*, $\sigma_{qq.} = 0$; details are available from the authors.

4. A simulation study

4.1 Design of the study

We now use a simulation study to evaluate the properties of our degrees-of-freedom estimators, and related variates, under moderate-sample-size conditions. We set up the simulation procedure as follows.

We considered four sets of V_h values from the NHANES III example for the Mexican-American subpopulation introduced in Section 1.1. Those four sets of V_h are the estimated $\hat{V}_h$ values from the variables BMPWT, HAR3, TCRESULT and HDRESULT, respectively, and are listed in Table 4.1. For each case, we used $(\beta_0, \beta_1) = (0, 1)$ and $\sigma_{qq} = 0$, in keeping with the results of Section 3, and thus $V_{Wh} = V_h$. Then, for each $h = 1, ..., L$, we obtained 10,000 realizations of the initial estimators $(\hat{Y}_{h1}, \hat{Y}_{h2}, \hat{V}_{Wh1}, \hat{V}_{Wh2})$ by assuming that the $\hat{Y}_{hi}$ are distributed as a normal random variable with mean zero and variance $2^{-1} V_h$; that $V_{Wh}^{-1}(m_{hi} - 1)\hat{V}_{Whi}$ is distributed as a chi-square random variable with $m_{hi} - 1$ degrees of freedom, where $m_{hi} = 11$ for all h and i; and the $\hat{Y}_{hi}$ and $\hat{V}_{Whi}$ are mutually independent. Note that in our data from NHANES III, the average number of secondary units for each PSU i in stratum h is about 11. For each replication, we computed $\hat{V}_h = (\hat{Y}_{h1} - \hat{Y}_{h2})^2$ and $\hat{V}_{Wh} = 2^{-1}(\hat{V}_{Wh1} + \hat{V}_{Wh2})$, and then carried out an errors-in-variables regression of $\hat{V}_h$ on $\hat{V}_{Wh}$ with measurement error variance $\hat{\sigma}_{uuh} = \widehat{\text{Var}}(\hat{V}_{Wh})$ using formula (2.2). This produced the coefficient estimators $(\hat{\beta}_0, \hat{\beta}_1)$, and the degrees-of-freedom estimators $\hat{d}_{mS}$ and $\hat{d}_{WS}$.

Table 4.1
"True" variances V_h used in simulation studies

Stratum	Case 1	Case 2	Case 3	Case 4
1	0.00E+00	0.00E+00	0.00E+00	0.00E+00
2	0.00E+00	0.00E+00	0.00E+00	0.00E+00
3	1.56E-04	7.67E-05	1.45E-02	1.76E-02
4	2.01E-04	3.57E-06	5.60E-02	4.55E-03
5	2.82E-04	4.88E-07	1.54E-03	2.91E-03
6	4.36E-04	0.00E+00	3.73E-03	8.60E-04
7	7.30E-04	2.14E-06	1.69E-02	1.13E-05
8	8.80E-04	1.30E-05	2.72E-02	1.40E-03
9	1.65E-03	1.16E-06	9.24E-03	1.35E-04
10	1.70E-03	9.46E-07	2.24E-03	1.77E-03
11	2.73E-03	0.00E+00	2.54E-04	1.32E-03
12	2.91E-03	5.40E-06	2.75E-02	6.40E-03
13	4.95E-03	3.73E-07	1.15E-02	5.38E-03
14	7.25E-03	2.90E-04	3.75E-02	6.97E-02
15	9.06E-03	9.81E-05	3.46E-01	7.58E-01
16	1.14E-02	7.47E-06	1.54E-02	4.75E-03
17	2.69E-02	9.65E-05	7.99E-02	1.01E-03
18	4.00E-02	1.12E-04	1.44E-01	1.77E-01
19	4.27E-02	2.68E-06	8.59E-02	3.88E-02
20	6.05E-02	7.57E-06	2.68E+00	7.18E-02
21	6.45E-02	1.17E-04	1.65E-01	4.52E-04
22	1.08E-01	1.05E-04	5.41E-01	1.98E-03

4.2 Coverage rates of *t*-based confidence intervals

For the four specified cases, Table 4.2 presents the simulated non-coverage probabilities obtained for t-based confidence intervals for the population mean $\bar{Y}$ that used the corresponding $\hat{d}$. For the severely heterogeneous cases (Cases 3 and 4), none of the degrees of freedom measures (not even the true d) leads to confidence intervals with coverage rates meeting the nominal rates $1 - \alpha$. That is, in extreme cases, the general Satterthwaite approach can be problematic for construction of confidence intervals, regardless of whether d, $\hat{d}_{mS}$, or $\hat{d}_{WS}$ is used to determine the t multiplier.

For Cases 1 and 2, the V_h values display less severe heterogeneity than in Cases 3 and 4. Table 4.2 shows that the simulated coverage probabilities with the true d for these two cases are slightly above 0.95. This overcoverage may be attributable to the fact that the variance estimator $\hat{V}(\hat{Y})$ is not distributed exactly as a multiple of a χ^2_d random variable, due to the heterogeneity of the V_h. Use of the standard degrees-of-freedom term $n - L$ or the modified estimator $\hat{d}_{mS}$ produces confidence intervals with coverage rates below the nominal level of 95%. On the other hand, use of our auxiliary-data-based term $\hat{d}_{WS}$ gives simulation based coverage rates close to the nominal 0.95 level.

Tables 4.3a and 4.3b display the empirical distributions of $\hat{d}$ and $2t_{\hat{d}}$ for the estimators $\hat{d}_{mS}$ and $\hat{d}_{WS}$. The simulated standard deviation of $t_{\hat{d}_{WS}}$ is smaller than that of $t_{\hat{d}_{mS}}$. In addition, the mean and median of $t_{\hat{d}_{WS}}$ are slightly larger than those of $t_{\hat{d}_{mS}}$. This is consistent with the undercoverage of the intervals based on $t_{\hat{d}_{mS}}$. Thus, under conditions similar to those for Cases 1 and 2 (or under conditions with less heterogeneity of V_h), it is worthwhile to consider the use of $\hat{d}_{WS}$ as a degrees-of-freedom estimator.

5. Application to a health survey

5.1 Preliminary model checks

We applied our proposed methods to the NHANES III data described in Section 1. It is important to check the modeling assumptions before we apply the proposed stability measures. First, for the Mexican-American sub-population described in Section 1, Table 5.1 gives values of $\hat{\kappa}_{xx}$ for the four variables which all have $\hat{\kappa}_{xx}$ values greater than 0.7.

Second, Figure 5.1 displays the scatter plots of $\hat{V}_h$ against $\hat{V}_{Wh}$ for the four variables with equal scales used for the horizontal and vertical axes. It shows that a linear relationship for the corresponding variables is plausible even if the relation would not be perfect and there are some outliers. Consequently, those four variables might be appropriate for the auxiliary-data-based method developed in Sections 2 and 3.

Table 4.2
Observed non-coverage rates for nominal 95% confidence intervals with $V_h = V_{Wh}$ in simulation study

	Case 1	Case 2	Case 3	Case 4
True d_S	6.26	6.04	2.38	2.20
Non-Coverage with t_{d_s}	0.0428	0.0443	0.0162	0.0164
Non-Coverage with t_{n-L}	0.0744	0.0788	0.1220	0.1263
Non-Coverage with $t_{\hat{d}_{mS}}$	0.0552	0.0567	0.0911	0.0905
Non-Coverage with $t_{\hat{d}_{WS}}$	0.0428	0.0466	0.0224	0.0220

Table 4.3a
Means and quantiles of degrees-of-freedom estimators $\hat{d}_{mS}$ and $\hat{d}_{WS}$: Cases 1 and 2

Cases	True d	Est.	[1]Mean $\hat{d}$	SD($\hat{d}$)	[2]Q(0.05)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.95)
1	6.26	$\hat{d}_{mS}$	9.33	3.33	4.45	6.86	9.01	11.41	15.30
		$\hat{d}_{WS}$	6.52	0.82	5.06	5.99	6.57	7.10	7.78
2	6.04	$\hat{d}_{mS}$	8.87	2.95	4.35	6.69	8.72	10.97	13.99
		$\hat{d}_{WS}$	6.34	0.96	4.67	5.69	6.42	7.06	7.80

[1] Mean denotes the average of the estimates, taken across all 10,000 replications.
[2] Q(·) indicates the quantile of the estimator, taken across all 10,000 replications.

Table 4.3b
Simulated non-coverage probabilities; and means and quantiles of *t*-multipliers for nominal 95% confidence intervals: Unequal true variances, cases 1 and 2

Cases	Est.	[1] $1-\hat{\alpha}$	[2] $M(2t_\alpha)$	$SD(2t_\alpha)$	[3]Q(0.05)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.95)
1	$\hat{d}_{mS}$	0.0552	4.62	0.36	4.26	4.38	4.52	4.75	5.37
	$\hat{d}_{WS}$	0.0428	4.83	0.16	4.64	4.72	4.80	4.90	5.13
	n - L	0.0744	4.15						
	True d_S	0.0428	4.85						
2	$\hat{d}_{mS}$	0.0567	4.66	0.36	4.29	4.41	4.55	4.78	5.41
	$\hat{d}_{WS}$	0.0466	4.87	0.21	4.64	4.72	4.83	4.97	5.28
	n - L	0.0788	4.15						
	True d_S	0.0443	4.89						

[1] $1-\hat{\alpha}$ is the simulated non-coverage probability of confidence intervals computed using estimated d.f.'s
[2] $M(2t_{0.975})$ is the average of twice of the 97.5% *t*-percentile value
[3] Q(.) indicates the quantile of $2t_{0.975,\,\hat{d}}$, taken across all replications.

Table 5.1
$\hat{\kappa}_{xx}$, estimates of model parameters, model diagnostics, and degrees of freedom estimates for four NHANES III variables (Mexican-American (Age 20-29) subgroup)

Variables	$\hat{\kappa}_{xx}$	$\tilde{\beta}_0$	se($\tilde{\beta}_0$)	$\tilde{\beta}_1$	se($\tilde{\beta}_1$)	Simulation based p-value for H_0: $\beta_0 = 0$	Simulation based p-value for H_0: $\beta_1 = 1$	$\hat{\tilde{\sigma}}_{qq.}$	$\hat{r}_{qq}$	$\hat{d}_{mS}$	$\hat{d}_{WS}$
BMPWT	0.75	-0.0013	0.0039	1.135	0.5429	0.3815	0.3541	-0.000	-0.43	15.49	10.04
HAR3	0.75	-0.000009	0.000012	1.095	0.3991	0.4229	0.3400	0.000	-0.83	14.94	8.30
TCRESULT	0.88	-0.146	0.0493	2.879	0.6252	0.0606	0.2259	-0.178	-0.77	5.88	6.59
HDRESULT	0.90	-0.042	0.0098	6.650	0.9988	<0.0001	0.1506	-0.017	-0.91	5.45	5.93

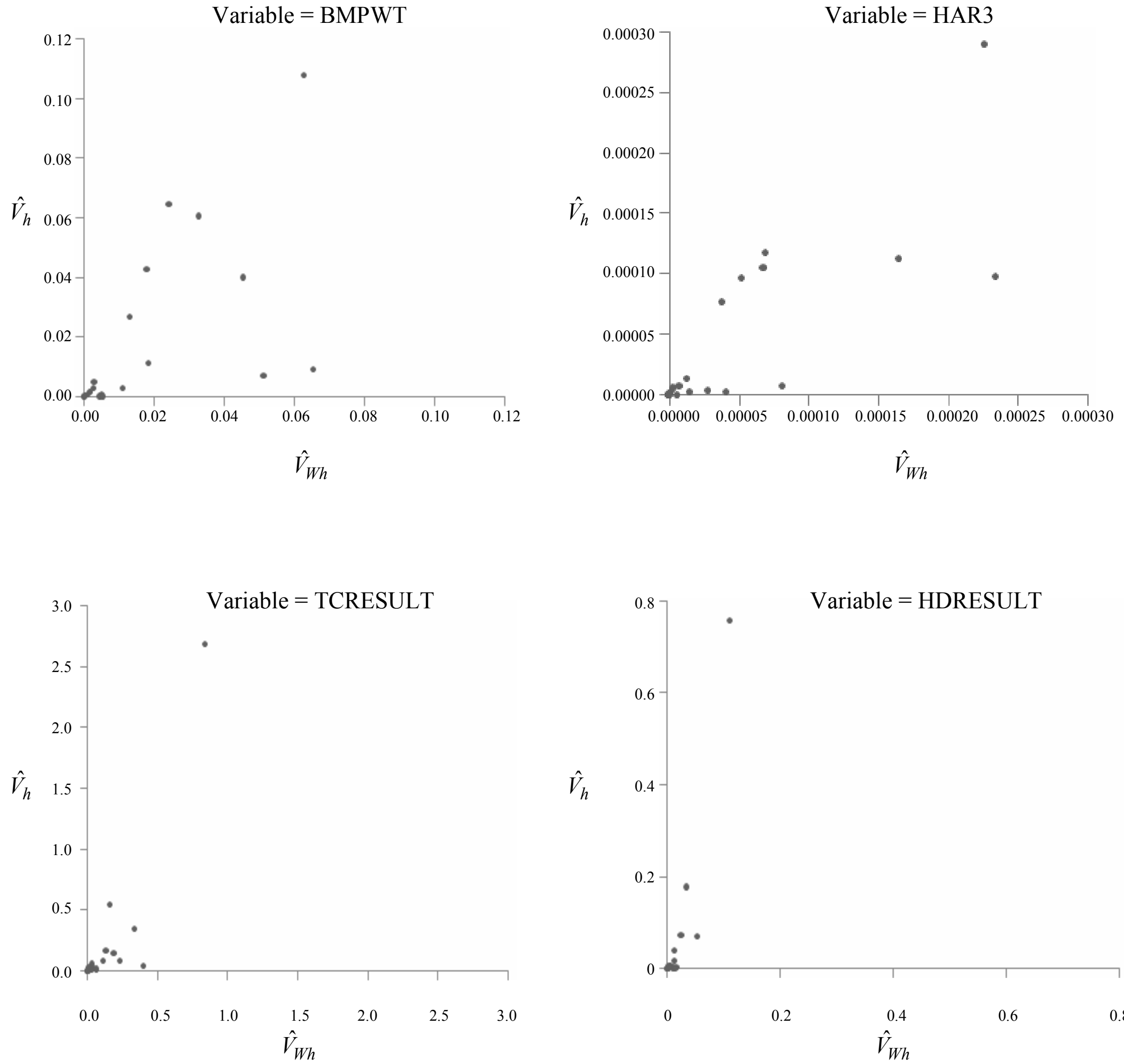

Figure 5.1 Plot of $\hat{V}_{Wh}$ vs. $\hat{V}_h$ for M-A (Age 20-29), Variable = BMPWT

5.2 An *ad hoc* test of $\bar{\sigma}_{qq\cdot} = 0$ under condition (C.1)

For all four variables considered in Table 5.1, the direct estimates $\hat{\hat{\sigma}}_{qq}$ of equation error variance (3.4) were negative or close to zero. That suggests that our χ^2-based estimator of σ_{eeh} as given in Section 3.1 might be too conservative or that $\bar{\sigma}_{qq\cdot}$ is indeed close to zero. This suggests that we need to re-examine the distributional assumption (C.1) in the NHANES III example. To do this, we considered the simulated distribution of $\hat{r}_{qq} \overset{\text{def}}{=} \hat{\hat{\sigma}}_{qq}/\hat{\sigma}_{ee}$, where division by $\hat{\sigma}_{ee}$ is used to avoid scale problems. The conditions and simulation design were as described in Section 4.1.

Table 5.2 reports results for $\hat{\sigma}_{ee}$ from expression (3.5), and $\hat{\hat{\sigma}}_{qq}$ computed from expression (3.4) with $\hat{\beta}_0$ set equal to zero and with $\hat{\beta}_1$, computed from expression (3.3). Table 5.2 reports the mean, standard deviation and selected quantiles of the simulated distribution of $\hat{r}_{qq\cdot}$ for the four variables. Table 5.3 reports the corresponding quantities for $\hat{r}_{qq}$, computed from $\hat{\hat{\sigma}}_{qq}$ given by expression (3.4) and with $\hat{\beta}_0$ and $\hat{\beta}_1$ computed from expression (3.3).

The results reported in Tables 5.2 and 5.3 lead to an ad hoc test of H_0: $\sigma_{qq} = 0$. Specifically, if the observed ratio $\hat{r}_{qq}$ falls above the upper 0.95 simulated quantile, then the assumption that $\bar{\sigma}_{qq} = 0$ may be problematic. Conversely, an observed $\hat{r}_{qq}$ below the .05 simulated quantiles in Tables 5.2 or 5.3 might indicate that $\hat{\sigma}_{eeh}$ is conservative, or may indicate violation of other parts of condition (C.1).

From Table 5.1, the values of $\hat{r}_{qq}$ for the variables are between -0.91 to -0.43. Except for HDRESULT, we do not have any strong evidence of violation of the model assumptions. However, for HDRESULT, the ratio $\hat{r}_{qq} = -0.91$ falls between the 0.01 and 0.05 quantiles reported in Table 5.2 and 5.3 for case 4. In general, values of $\hat{r}_{qq}$ that fall above the 0.95 or 0.99 quantiles of Tables 5.2 or 5.3 would be consistent with values of $\bar{\sigma}_{qq\cdot}$ greater than zero. The observed value $\hat{r}_{qq} = -0.91$ is not necessarily consistent with $\bar{\sigma}_{qq\cdot} > 0$, but may indicate violation of one or more conditions in (C.1)-(C.4).

Table 5.2
Means and quantiles of $\hat{r}_{qq} = \hat{\sigma}_{ee}^{-1}\hat{\sigma}_{qq.}$ ($\hat{\beta}_0 = 0$)

Cases	1M($\hat{r}_{qq}$)	SD($\hat{r}_{qq}$)	2Q(0.01)	Q(0.05)	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	Q(0.95)	Q(0.99)
1	-0.50	0.66	-1.71	-1.30	-1.15	-0.99	-0.79	0.16	0.54	0.60	0.65
2	-0.48	0.68	-1.72	-1.32	-1.16	-0.99	-0.76	0.23	0.57	0.62	0.66
3	-0.19	0.42	-1.01	-0.84	-0.74	-0.53	-0.20	0.17	0.38	0.46	0.55
4	-0.20	0.39	-1.00	-0.82	-0.72	-0.51	-0.20	0.11	0.34	0.44	0.56

[1] M denotes the average of the estimates, taken across all 10,000 replications.
[2] Q(.) indicates the quantile of the estimator, taken across all 10,000 replications.

Table 5.3
Means and quantiles of $\hat{r}_{qq} = \hat{\sigma}_{ee}^{-1}\hat{\bar{\sigma}}_{qq.}$

Cases	1M($\hat{r}_{qq}$)	SD($\hat{r}_{qq}$)	2Q(0.01)	Q(0.05)	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	Q(0.95)	Q(0.99)
1	-0.56	0.62	-1.85	-1.34	-1.17	-1.00	-0.80	0.05	0.38	0.44	0.52
2	-0.56	0.62	-1.91	-1.37	-1.18	-1.00	-0.78	0.06	0.35	0.42	0.50
3	-0.24	0.42	-1.16	-0.90	-0.79	-0.57	-0.22	0.12	0.29	0.36	0.45
4	-0.24	0.38	-1.09	-0.87	-0.75	-0.53	-0.22	0.06	0.25	0.33	0.44

[1] M denotes the average of the estimates, taken across all 10,000 replications.
[2] Q(.) indicates the quantile of the estimator, taken across all 10,000 replications.

5.3 Coefficient estimates and degrees-of-freedom estimates

Because our data were consistent with $\bar{\sigma}_{qq.} = 0$ for all four cases, we used the methods of Fuller (1987, page 124) to produce estimates of β_0 and β_1 appropriate for a model (3.1) – (3.2) with no equation error; details are available from the authors. Table 5.1 also reports the resulting coefficient estimates $\tilde{\beta}_0$ and $\tilde{\beta}_1$, and their standard errors, $se(\tilde{\beta}_0)$ and $se(\tilde{\beta}_1)$. Recall from Section 3.1 that under model (3.1) – (3.2), if $\beta_0 = 0$ and $\beta_1 \neq 0$, then each stratum variance V_h is a constant multiple of the within-PSU variance V_{Wh}, and $\hat{d}_{WS}$ in (1.8) may be an appropriate estimator of d. Section 5.2 already considered the condition $\bar{\sigma}_{qq.} = 0$. To test the null hypothesis H_0: $\beta_0 = 0$, we use the test statistic, $t_0 = \tilde{\beta}_0 / se(\tilde{\beta}_0)$. In some practical errors-in-variables work, quantities like t_0 are compared with a standard normal or t reference distribution. However, simulation work based on the four cases from Section 4.1 indicated that the null distribution of t_0 deviated substantially from these customary reference distributions. This is due to the very skewed distributions of the response variables $\hat{V}_h$ used in the errors-in-variables regression. Consequently, we used standard methods to develop a simulation-based reference distribution for t_0. Column 7 of Table 5.1 reports the resulting left-tailed p-value. (Due to negative point estimates $\tilde{\beta}_0$, we have chosen to report the left-tailed p-values here. In other cases, it may be of interest to report right-tailed or two-tailed p-values for β_0). There is strong evidence against H_0: $\beta_0 = 0$ for the variable HDRESULT, and the moderate evidence against H_0: $\beta_0 = 0$ for TCRESULT. Thus, it may not be appropriate to use $\hat{d}_{WS}$ for these two variables. Now consider the slope coefficient β_1, and suppose that $\sigma_{qqh} = 0$ so $q_h = 0$ with probability one. Then expressions (1.5) and (3.1), and the nonnegativity of V_{Bh} implies that $0 \le V_{Bh} = V_h - V_{Wh} = \beta_0 + (\beta_1 - 1)V_{Wh}$. Consequently, if $\beta_0 = 0$, then $\beta_1 \ge 1$ and $\beta_1 = 1$ is equivalent to $V_h = V_{Wh}$. This final condition is of practical interest because some authors have noted cases in which V_{Bh} is small relative to V_{Wh}, or equivalently, $V_h \doteq V_{Wh}$. See for example, Wolter (1985, page 46). To test H_0: $\beta_1 = 1$ against the one-sided alternative H_1: $\beta_1 > 1$, we used the statistic $t_1 = (\tilde{\beta}_1 - 1)/se(\tilde{\beta}_1)$. For reasons similar to those for t_0, we developed simulation-based reference distributions for t_1 under each of Cases 1 through 4. Column 8 of Table 5.1 reports the resulting one-tailed p-values.

The last two columns of Table 5.1 report the degree-of-freedom estimators $\hat{d}_{mS}$ and $\hat{d}_{WS}$. For HAR3 and BMPWT, $\hat{d}_{mS}$ gives substantially larger values than $\hat{d}_{WS}$.

6. Discussion

This paper has considered estimation of a degrees-of-freedom term d used to quantify the variability of a standard design based variance estimator $\hat{V}(\hat{Y})$. The fundamental issue is that under a design involving heterogeneous stratum-level variances and small numbers of primary sample units selected per stratum, the Satterthwaite-type estimator $\hat{d}_{mS}$ may perform poorly. We developed an alternative estimator $\hat{d}_{WS}$ based on within-primary-sample unit variance estimators $\hat{V}_{Wh}$. This alternative estimator is a solution to an unbiased estimating equation (1.1) for d, provided the proportionality condition (1.6) is satisfied. Also, the variance of the approximate distribution of $\hat{d}_{WS}$ is smaller than that of $\hat{d}_{mS}$, provided the number of secondary sample units selected within each primary unit is large, in the sense defined by Result 2.2.

Section 3 developed errors-in-variables methods for testing the adequacy of the proportionality condition (1.6), and suggested some related diagnostics. The simulation study in Section 4, in conjunction with the data analysis in Section 5, indicated that under moderate amounts of heterogeneity, $\hat{d}_{WS}$ can perform better than $\hat{d}_{mS}$, in terms of the distributional properties of these estimators of d, and in terms of the coverage rates and widths of associated confidence intervals for the population totals Y. However, as one would expect from standard large-sample theory, neither estimator performs well under severe heterogeneity.

One could in principle consider use of the errors-in-variables estimators $(\hat{\beta}_0, \hat{\beta}_1, \hat{\bar{\sigma}}_{qq.})$, in conjunction with the $\hat{V}_h$ and $\hat{V}_{Wh}$, to construct an alternative estimator of d that will be consistent under the general errors-in-variables model (3.1) - (3.2), and will not require the restrictive condition (1.6). However, simulation results in Jang (1996) indicated that the resulting estimator $\hat{d}_{EIV}$, say, did not perform well under the design conditions used in Section 5.

The principal results of Sections 1 through 3 extend readily from the within-primary-unit variances V_{Wh} to more general auxiliary variables X_h. For such extensions, the principal issues remain the adequacy of the proportionality approximation (1.6); and the amount of sampling error in the auxiliary estimators $\hat{X}_h$, say, relative to the error in the basic stratum-level variance estimator $\hat{V}_h$.

Acknowledgements

The authors thank the U.S. National Center for Health Statistics for providing access to the NHANES III dataset, and thank V.L. Parsons, C. Johnson and L.R. Curtin for sharing a wealth of information regarding the NHANES III. This research was supported in part by the U.S. National Center for Health Statistics. The views expressed in this paper are those of the authors and do not necessarily represent the policies of the U.S. National Center for Health Statistics or the U.S. Bureau of Labor Statistics.

Appendix A

Proof of result 2.2

Consider a nonlinear function $B^{-1}A^2$ of two estimators A and B with means μ_A and μ_B, respectively. Then, the variance of the leading term of a Taylor expansion of $B^{-1}A^2$ is

$$\frac{4\mu_A^2}{\mu_B^2}\operatorname{Var}(A) - 4\frac{\mu_A^3}{\mu_B^3}\operatorname{Cov}(A,B) + \frac{\mu_A^4}{\mu_B^4}\operatorname{Var}(B) \text{ .(A.1)}$$

Now we define the following two estimators: $L^{-1}\hat{d}_{S1} = B_1^{-1}A_1^2$ and $L^{-1}\hat{d}_{S2} = B_2^{-1}A_2^2$, where $A_1 = L^{-1}\sum_{h=1}^L \hat{V}_h$, $B_1 = L^{-1}\sum_{h=1}^L \hat{V}_h^2$, $A_2 = L^{-1}\sum_{h=1}^L \hat{V}_{Wh}$, and $B_2 = L^{-1}\sum_{h=1}^L \{\hat{V}_{Wh}^2 - \widehat{\operatorname{Var}}(\hat{V}_{Wh})\}$.

Assume conditions (C.1), (C.2) and (C.3). In addition, define $\hat{F}_{L\hat{d}_{S1}}$ and $\hat{F}_{L\hat{d}_{S2}}$ to be the leading terms of Taylor expansions of $L^{-1}\hat{d}_{S1} - \mu_{B_1}^{-1}\mu_{A_1}^2$ and $L^{-1}\hat{d}_{S2} - \mu_{B_2}^{-1}\mu_{A_2}^2$, respectively. Also, recall that if D is distributed as a chi-square random variable on d degrees of freedom, then $V(D) = 2d$, $E(D^3) = d(d+2)(d+4)$, and $V(D^2) = 8d(d+2)(d+3)$. Then the corresponding components of $\operatorname{Var}(\hat{F}_{L\hat{d}_{S1}})$ and $\operatorname{Var}(\hat{F}_{L\hat{d}_{S2}})$ in (A.1) are

$$\operatorname{Var}(A_1) = 2L^{-2}\sum_{h=1}^L V_h^2,$$

$$\operatorname{Var}(A_2) = 2(m_0-1)^{-1}L^{-2}\sum_{h=1}^L V_{Wh}^2$$

$$\operatorname{Var}(B_1) = 96L^{-2}\sum_{h=1}^L V_h^4,$$

$$\operatorname{Var}(B_2) = 8(m_0-1)^{-2}(m_0+1)L^{-2}\sum_{h=1}^L V_{Wh}^4$$

$$\operatorname{Cov}(A_1,B_1) = 12L^{-2}\sum_{h=1}^L V_h^3,$$

and (A.2)

$$\operatorname{Cov}(A_2,B_2) = 4(m_0-1)^{-1}L^{-2}\sum_{h=1}^L V_{Wh}^3.$$

Since we assume $n_h = 2$ and $m_h = m_0$ for all $h = 1, 2, \ldots, L$, we have

$$L^{-1}\hat{d}_{mS} = L^{-1}(3L+14)^{-1}(9L)\hat{d}_{S1} \tag{A.3}$$

and

$$L^{-1}\hat{d}_{WS} = L^{-1}\hat{d}_{S2}\,. \tag{A.4}$$

Under condition (1.6), $\mu_{A1} = \beta_1 \mu_{A2}$,

$$\begin{aligned}\mu_{B1} &= 3\beta_1^2 \mu_{B2},\\ \mathrm{Var}(A_1) &= (m_0 - 1)\beta_1^2 \mathrm{Var}(A_2),\\ \mathrm{Var}(B_1) &= 12(m_0+1)^{-1}(m_0-1)^2 \beta_1^4 \mathrm{Var}(B_2)\end{aligned}$$

and

$$\mathrm{Cov}(A_1, B_1) = 3(m_0 - 1)\beta_1^3 \mathrm{Cov}(A_2, B_2) \qquad \text{(A.5)}$$

Substituting (A.5) into (A.1) leads to,

$$\begin{aligned}\mathrm{Var}(\hat{F}_{L\hat{d}_{S1}}) &= \frac{4}{9}(m_0 - 1)\frac{\mu_{A_2}^2}{\mu_{B_2}^2}\mathrm{Var}(A_2)\\ &- \frac{4}{9}\frac{\mu_{A_2}^3}{\mu_{B_2}^3}(m_0 - 1)\mathrm{Cov}(A_2, B_2)\\ &+ \frac{4(m_0-1)^2}{27(m_0+1)}\frac{\mu_{A_2}^4}{\mu_{B_2}^4}\mathrm{Var}(B_2)\\ &= \frac{1}{9}(m_0 - 1)a - \frac{1}{9}(m_0 - 1)b + \frac{4(m_0-1)^2}{27(m_0+2)}c \qquad \text{(A.6)}\end{aligned}$$

where $\mathrm{Var}(L^{-1}\hat{d}_{WS}) = a - b + c$. With large $L, \mathrm{Var}(\hat{\mathrm{F}}_{L\hat{d}_{mS}}) = (m_0 - 1)a - (m_0 - 1)b + \{3(m_0+2)\}^{-1} 4(m_0-1)^2 c$. Thus for large L, $V(\hat{F}_{L\hat{d}_{mS}}) - V(\hat{F}_{L\hat{d}_{WS}}) \doteq (m_0 - 2)a - (m_0 - 2)b + \{3(m_0+2)\}^{-1}(4m_0^2 - 11m_0 - 2)c$. Therefore, $\lim_{L\to\infty} V_{Lm} - \lim_{L\to\infty} V_{LW} \geq 0$ if $m_0 \geq \lim_{L\to\infty}\{2(3a - 3b + 4c)\}^{-1}\{11c + \sqrt{144a^2 + 144b^2 + 153c^2 - 288ab + 216ac - 216bc}\}$. In particular, $\lim_{L\to\infty} V_{Lm} - \lim_{L\to\infty} V_{LW}$ becomes greater than or equal to zero when $m_0 = 10$ regardless of values of a, b, and c. Because it is an increasing function in m_0, for all values of $m_0 \geq 10$, $\lim_{L\to\infty} V_{Lm} \geq \lim_{L\to\infty} V_{LW}$.

References

Casady, R., Dorfman, A.H. and Wang, S. (1998). Confidence intervals for sub-domain parameters when the sub-domain sample size is random. *Survey Methodology*, 24, 57-67.

Cochran, G.C. (1977). *Sampling Techniques*, 3rd Ed. New York: John Wiley & Sons, Inc.

Eltinge, J.L., and Jang, D. (1996). Stability measures for variance component estimators under a stratified multistage design. *Survey Methodology*, 22, 157-165.

Fuller, W.A. (1987). *Measurement Error Models*. New York: John Wiley & Sons, Inc.

Isaki, C.T. (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78, 117-123.

Jang, D. (1996). *Stability of Variance Estimators Under Complex Sampling Designs*. Unpublished Ph.D. dissertation, Department of Statistics, Texas A&M University, College Station, Texas.

Kendall, M., Stuart, A. and Ord, J.K. (1983). *The Advanced Theory of Statistics*, *Volume* 3: *Design and Analysis*, *and Time Series*. New York: Macmillan.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jacknife and balanced repeated replication methods. *Ann. Statist*., 9, 1010-1019.

National Center for Health Statistics (1996). NHANES III Reference Manuals and Reports, CD-ROM GPO, 017-022-1358-4. Washington, D.C.: United States Government Printing Office.

Satterthwaite, F.E. (1941). Synthesis of variance. *Psychometrika*, 6, 309-316.

Satterthwaite, F. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110-114.

Smith, H.F. (1936). The problem of comparing the results of two experiments with unequal errors. *Journal of the Council for Scientific and Industrial Research*, 9, 211-212.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.