The Method of Laplace and BRR: A Hybrid Variance Estimation Method in Surveys

Yves Thibaudeau, Eric Slud

Yves Thibaudeau, U.S. Census Bureau, Washington, DC 20233, Yves.Thibaudeau@Census.gov

1. Introduction[1]

We present hybrid estimators for population totals subject to transitions through time, along with a corresponding hybrid variance estimator. Table 1.1 classifies selected estimators according to their type of methodology (design based, model-based) and their fundamental perspective (frequentist, Bayesian). The estimator of population totals we are proposing is a hybrid estimator in the sense that it involves a component akin the Horvitz-Thompson estimator, a design-based estimator, and another component defined through Maximum Likelihood estimation.

Table 1.2 classifies variance estimators along the same two axes. In the paper we also present a hybrid variance estimator. This variance estimator is a hybrid because it integrates balanced repeated replication (BRR), a frequentist procedure, and the Method of Laplace, a Bayesian procedure.

Ultimately we will show that, although the method of Laplace originates in the Bayesian perspective, it has properties that allow using it in the frequentist perspective as well. Furthermore, it can be part of a diagnostic tool kit to validate models in model-based estimation. This is important, because as we will see, structurally model-based estimator are advantageous over estimator using models just to patch-up missing data, if the model is correct.

According to our definition, hybrid estimators only involve methods which truly combine estimators at both end of one axis in table 1 or 2. These should be distinguished from "bridge" estimators. For example, multiple imputation (MI) was originally proposed by Rubin (1977, 1987 p. 116) as a randomization-valid, by extension design-based, method. But later, MI became popular as a model-based method. So MI is a bridge estimator. Tables 1.1 and 1.2 are by no mean exhaustive or definitive classifications. They are meant as a reference point to frame our research.

In section 2 we describe the example we use in the paper to introduce our hybrid estimators: The Survey of Income and Program Participation. Section 3 introduces basic hybrid estimators in presence of missing data. Section 4 recalls BRR, and expands our hybrid variance estimators involving the method of Laplace for variance estimation. Section 5 gives a frequentist validation of both our hybrid estimators in terms of their interpretations relative to population statistics. Section 6 concludes by giving an appreciation of the potential of the method of Laplace as a diagnostic tool.

Table 1.1 – Selected Estimation Techniques

Perspective -------------------- *Methodology*	*Frequentist*	*Bayesian*
Design-Based	Horvitz-Thompson, WMLE	Multiple-Imputation
Model-Based	GREG, MLE	MLE, Multiple-Imputation

Table 1.2 – Selected Variance Estimation Techniques

Perspective -------------------- *Methodology*	*Frequentist*	*Bayesian*
Design-Based	Expansion Estimator, BRR, Bootstrap, Jackknife	Multiple Imputation
Model-Based	Linearization, BRR, Bootstrap, Jackknife	Bootstrap, Multiple-Imputation, Method of Laplace

[1] *This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress). Any views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not necessarily those of the U.S. Census Bureau*

2. Example: The survey of Income and Program Participation

For this expose we assume a nontrivial, but simple, structure to illustrate the problem of estimating population totals and their variances. We use the survey of income and program participation (SIPP), a longitudinal survey conducted by the U.S. Census Bureau, as background for our example.

SIPP measures the economic well being of the U.S. general population in relation to participation to government social programs, such as unemployment compensations and several programs of economic assistance. Participants to the survey are asked to report on their status in relation to several programs for each month they are in the cohort. But participants are interviewed only every four months in "waves" installments. Missing data are endemic to SIPP. A frequent situation is that of an initial full response to the survey, at the first wave, but only fragmented responses at subsequent waves.

Table 2.1 shows sample counts for the state of California for wave 1 and wave 2 of the 2004 "panel. It gives a cross-classification of each screened respondent by the answers to two labor-force questions: "Were you looking for a job on the last month of wave 1 (2)?" and "Were you on layoff anytime during the last month of wave 1 (2)?" Tables 2.2, 2.3, 2.4 give counts involving missing responses at wave 2. For simplicity of the expose, the cases involving missing response at wave 1 were discarded. But, in general, the techniques we present recuperate these cases.

Table 2.1 – Classification for Units with Complete Wave 1 and Wave 2 Information

			Wave 2			
			Not Looking for Work		Looking for Work	
			Not Layoff	Layoff	Not Layoff	Layoff
Wave 1	Not Looking	Not on layoff	109	1	28	4
		On Layoff	0	6	3	7
	Looking	Not on Layoff	37	2	36	3
		On Layoff	7	10	3	14

Table 2.2 – Classification for Units with Partial Wave 2 Information

			Wave 2	
			Not Looking for Work	Looking for Work
Wave 1	Not Looking.	Not Layoff	533	134

Table 2.3 – Classification for Units with Partial Wave 2 Information (cont'd)

			Wave 2	
			Not on Layoff	Layoff
Wave 1	Not Looking	Not Layoff	3	0

Table 2.4 – Classification for units with No Wave 2 Information

Wave 1	Not Looking	Not on Layoff	121

3. Estimation of Population Totals

For a given unit in the survey, let *I*, *J* represent respectively the responses to the "looking for work" and "on layoff" questions at wave 1, and let *K*, *L* represent respectively the responses to the "looking for work" and "on layoff" questions at wave 2. Define

$$I, J, K, L \quad = \quad \begin{cases} 1 & \text{if the correct answer is no} \\ 2 & \text{if the correct answer is yes} \end{cases} \tag{1}$$

In general, we represent the Horvitz-Thompson weighted total estimator of the population total in "state" (*I, J*) at wave 1, and state (*K, L*) at wave 2 by

$$T^{W}_{ijkl} \;=\; \sum_{s\in S} W_s \mathrm{I}_s \left(I=i,\, J=j,\, K=k,\, L=l\right) \tag{2}$$

Where W_s is the weight of unit *s*, indexed over the entire sample *S*. The "+" sign in lieu of the index indicates a marginal summation over that index. Because *K* or *L* can be missing, the following partial totals are larger than 0.

$$T^{W}_{ijk(l)} \;=\; \sum_{s\in S} W_s \mathrm{I}_s \left(I=i,\, J=j,\, K=k;\, L \text{ missing}\right) \quad ; \quad T^{W}_{ij(k)l} \;=\; \sum_{s\in S} W_s \mathrm{I}_s \left(I=i,\, J=j,\, L=l;\, K \text{ missing}\right)$$

$$T^{W}_{ij(k)(l)} \;=\; \sum_{s\in S} W_s \mathrm{I}_s \left(I=i,\, J=j;\, K, L \text{ missing}\right) \tag{3}$$

These partial sample totals will be integrated to estimators to predict population totals.

3.1 The Horvitz Thompson Estimator when no Data Are Missing

For purpose of illustration, the paper focuses on estimating the population total of the "on layoff" at wave 2. So this target population is defined by $L = 2$. When no data are missing, we represent the traditional Horvitz-Thompson weighted-total estimator for the size of the "on layoff" population by $\hat{N}_{+++2}$, where

$$\hat{N}_{+++2} \;=\; T^{W}_{+++2} \;=\; \sum_{s\in S} W_s \mathrm{I}_s \left(L=2\right) \tag{4}$$

3.2 A model for Transition Probabilities

We posit a simple loglinear model to describe the transition process between waves. Let π_{ijkl} be the inclusion probability for those units who answered (*i, j*) to questions (*I, J*), and who answered (*k, l*) to questions (*K, L*). So we have $\pi_{ijkl} = \Pr\left\{\left(I, J, K, L\right) = \left(i, j, k, l\right)\right\}$. Define the transition probability P_{ijkl} from wave 1 to wave 2 as follow:

$$P_{ijkl} = \Pr\left\{\left(K, L\right) = \left(k, l\right) \middle| \left(I, J\right) = \left(i, j\right)\right\} \qquad = \frac{\pi_{ijkl}}{\sum\limits_{m,n=1}^{2} \pi_{ijmn}} \tag{5}$$

Where $\sum_{k,l=1}^{2} P_{i\,j\,k\,l} = 1; \quad i, j = 1, 2$. So, (I, J), the answers to the wave 1 questions, conceptually play the role of post-stratifying variables. We define a loglinear model for the set of transition probabilities $\left\{P_{i\,j\,k\,l}\right\}_{i,j,k,l=1}^{2}$ implicitly, by specifying a classic loglinear model on $\left\{\pi_{i\,j\,k\,l}\right\}_{i,j,k,l=1}^{2}$, the set of inclusion probabilities (Bishop Fienberg Holland p. 263). The model is a hierarchical log linear model with no interaction effects higher than second order. Additionally, some second order interaction effects are suppressed in order to impose a conditional independence structure between waves and questions. Specifically, the interaction between questions I and L, and the interaction between K and J are suppressed. This means that, given "layoff at wave1" and "looking for work at wave 2" is known, then "layoff at wave 2" and "looking for work at wave 1" are independent. Similarly, given "looking for work at wave 1" and "layoff at wave 2" is known, then "layoff at wave 1" and "looking for work at wave 2" are independent. This model was fitted to the data presented in tables 2.1 – 2.4 and exhibits a satisfactory fit. The suppressed 2nd order interactions turned out indeed not to be significant.

Given this structure, the likelihood of $\left\{\pi_{i\,j\,k\,l}\right\}_{i,j,k,l=1}^{2}$ factors into two separate likelihoods for $\left\{P_{i\,j\,k\,l}\right\}_{i,j,k,l=1}^{2}$ and $\left\{\pi_{i\,j\bullet\bullet}\right\}_{i,j}^{2}$, respectively (Bishop Fienberg Holland p 263). The later is the set of joint marginal inclusion probabilities for I and J. Since the original model on the full inclusion probabilities had eight degrees of freedom and the joint inclusion probabilities have three degrees of freedom, that leaves five degrees of freedom for the conditional probability structure, which is ultimately the object of the modeling here. Pfefferman, Skinner and Humphreys (1998) propose more complex models for transition probabilities in the context of a study on measurement error. We will use this model to support the estimation of population totals: First we use the model only to "patch up" the missing data, assuming a missing at random mechanism. Second, we make a more structural use of the model by making it an integral part of the estimation process, rather than using it only as an auxiliary for adjusting for nonresponse.

3.3 The Horvitz-Thompson Estimator Adjusted for Missing Data

Recall, $\hat{N}_{+++2}$ in (4) covers only units providing a value for L, "on layoff at wave2". So $\hat{N}_{+++2}$ skips several units because of missing data, as suggested by tables 2.2 – 2.4. To remedy this problem, we introduce an adjusted H-T estimator. The adjusted H-T estimator is the sum of the weighted totals for the observed cases of "on layoff" at wave 2, and prorated components based on complete observations at wave 1 and partial observations at wave 2. The underlying assumption for the missing data mechanism is that of observations missing at random (MAR) between waves. We get:

$$\hat{N}^{a}_{+++2} = T^{W}_{111(l)} \left(\frac{\hat{P}_{1112}}{\hat{P}_{1111} + \hat{P}_{1112}} \right) + T^{W}_{112(l)} \left(\frac{\hat{P}_{1122}}{\hat{P}_{1121} + \hat{P}_{1122}} \right) + T^{W}_{11(k)(l)} \left(\hat{P}_{1112} + \hat{P}_{1122} \right) + T^{W}_{+++2} \qquad (6)$$

3.5 The Forecast Estimator Adjusted for Missing Data

We now present an estimator that uses the model in a much more structural manner than the estimator in (6). We define

$$\begin{aligned} \hat{\hat{N}}^{a}_{+++2} = \; & T^{W}_{111(l)} \left(\frac{\hat{P}_{1112}}{\hat{P}_{1111} + \hat{P}_{1112}} \right) + T^{W}_{112(l)} \left(\frac{\hat{P}_{1122}}{\hat{P}_{1121} + \hat{P}_{1122}} \right) + T^{W}_{11(k)(l)} \left(\hat{P}_{1112} + \hat{P}_{1122} \right) \\ & + \left(T^{W}_{11++} - T^{W}_{111(l)} - T^{W}_{112(l)} - T^{W}_{11(k)(l)} \right) \left(\hat{P}_{1112} + \hat{P}_{1122} \right) \qquad (7) \\ & + T^{W}_{12++} \left(\hat{P}_{1212} + \hat{P}_{1222} \right) + T^{W}_{21++} \left(\hat{P}_{2112} + \hat{P}_{2122} \right) + T^{W}_{22++} \left(\hat{P}_{2212} + \hat{P}_{2222} \right) \end{aligned}$$

The first three terms on the RHS of (7) are the same nonresponse-adjustment terms as in (6). The last four terms replace the weighted sum T^{W}_{+++2} in (6), which was based only on wave 2 information. Because (7) makes use of all the information, both from wave 1 and wave 2, we conjecture gains in efficiency, if the model is correct.

4. Variance Estimation

4.1 Balanced Repeated Replication

Wolter (1985 p 110) describes variance estimation based on balanced half-sample, which is the basis for the balanced repeated replication (BRR) technique to estimate variance of linear estimator. Judkins (1990) describes "Fay's method," for BRR, which makes use of all the observations in the sample to construct individual replicates. At first, we use Fay's method to evaluate variances. In the SIPP example for the state of California there are 23 dual-PSU strata defined for variance estimation. We use a Hadamard matrix of dimension 24 x 24, which leads to the construction of 24 replicates in implementing BRR.

Fay's method based on the observations in tables 2.1-2.4. leads to the following estimates for the variances of $\hat{N}^{a}_{++2+}$ in (6) and $\hat{\hat{N}}^{a}_{+++2}$ in (7) :

$$\hat{V}^{BRR}\left[\hat{N}^{a}_{+++2}\right] = 3.76 \times 10^{9} \quad ; \quad \hat{V}^{BRR}\left[\hat{\hat{N}}^{a}_{+++2}\right] = 2.94 \times 10^{9} \qquad (8)$$

As expected, the variance of $\hat{\hat{N}}^{a}_{+++2}$ is substantially less than that of $\hat{N}^{a}_{+++2}$. We will attempt to validate our model, which is fundamental to the validity of $\hat{\hat{N}}^{a}_{+++2}$ itself, in later sections.

4.2 Variance Estimation through Linearization

$\hat{\hat{N}}^{a}_{+++2}$ in (7) can be linearized into the sum of a model-based component, generated by the MLE $\left\{\hat{P}_{kli\,j}\right\}^{2}_{i,j,k,l=1}$, and a design-based component, generated by the H-T vector of weighted totals $\boldsymbol{T} = \left(T^{W}_{111(l)}, T^{W}_{112(l)}, T^{W}_{11(k)(l)}, T^{W}_{11++}, T^{W}_{12++}, T^{W}_{21++}, T^{W}_{22++}\right)^{t}$. In our situation, the two components of the linearization are uncorrelated. That is $\boldsymbol{T}$ does not convey any information on how wave 1 weighted totals are allocated between the "not on layoff" and "on layoff" subcategories at wave 2. Now consider the standard variance decomposition:

$$V\left[\hat{\hat{N}}^{a}_{+++2}\right] = E\left[V\left[\hat{\hat{N}}^{a}_{+++2} \middle| \boldsymbol{T}\right]\right] + V\left[E\left[\hat{\hat{N}}^{a}_{+++2} \middle| \boldsymbol{T}\right]\right] \qquad (9)$$

Given the design-based and model-based components of the linearization are uncorrelated, the two terms on the RHS of (9) correspond precisely to the contribution to the total variance of $\hat{\hat{N}}^{a}_{+++2}$ of the model-based and design-based components, in that order.

4.3 The Method of Laplace for Posterior Variances: A Large Sample approximation

We can use the method of Laplace to evaluate the contribution from the model-based component of a linearized estimator. In general, let $M\left(\hat{\boldsymbol{\lambda}}\ ;\ \boldsymbol{T}\right)$ be a sample statistic that is strictly a function of $\hat{\boldsymbol{\lambda}}$, the MLE of $\boldsymbol{\lambda}$, and of a vector of weighted totals $\boldsymbol{T}$. For instance, later we will set $M\left(\hat{\boldsymbol{\lambda}}\ ;\ \boldsymbol{T}\right) = \hat{\hat{N}}^{a}_{+++2}$. With no loss of generality, we define $\hat{\boldsymbol{\lambda}} = \max_{\boldsymbol{\lambda}} \arg\left[L\left(\boldsymbol{\lambda}\ ;\ (\boldsymbol{T},\boldsymbol{X})\right)\right]$, where $L\left(\boldsymbol{\lambda}\ ;\ (\boldsymbol{T},\boldsymbol{X})\right)$ is the likelihood of $\boldsymbol{\lambda}$, and $\boldsymbol{X}$ augments $\boldsymbol{T}$ to form the sufficient statistics $(\boldsymbol{T}, \boldsymbol{X})$ for $\{P_{ijkl}\}_{i,j,k,l=1}^{2}$ .Next, define $\hat{\lambda}_c = \max_{\lambda} \arg\left[L(\lambda\ ;\ (\boldsymbol{T}, \boldsymbol{X}))(M(\lambda\ ;\ \boldsymbol{T}) + c)\right]$. We propose using the low-order version of the Laplace large-sample approximation proposed by Tierney, Kass, and Kadane (1989), to approximate the first term on the RHS of (9), in effect the "posterior variance". We have

$$\hat{E}\left[V\left[\hat{M}\left(\hat{\lambda}\ ;\ \boldsymbol{T}\right)\middle|\boldsymbol{T}\right]\right] = \hat{V}^{L}\left[\hat{M}\left(\hat{\lambda}\ ;\ \boldsymbol{T}\right)\middle|\boldsymbol{T}\right] = \lim_{c\to\infty} c\left|M\left(\hat{\lambda}_c\ ;\ \boldsymbol{T}\right) - M\left(\hat{\lambda}\ ;\ \boldsymbol{T}\right)\right| \qquad \textbf{(10)}$$

Applying (10) on the data in tables 2.1 – 2.4 we obtain

$$\hat{E}\left[V\left[\hat{\hat{N}}^{a}_{+++2}\middle|\boldsymbol{T}\right]\right] = \hat{V}^{L}\left[\hat{\hat{N}}^{a}_{+++2}\middle|\boldsymbol{T}\right] = 2.22 \times 10^{9} \qquad \textbf{(11)}$$

In the context of the frequentist linearization of $\hat{\hat{N}}^{a}_{+++2}$, the interpretation for (11) is simply that of a large-sample approximation. The second variance term on the RHS of (9) corresponds to the contribution of the design-based component of the linearization to the total variance of $\hat{\hat{N}}^{a}_{+++2}$. We use Fay's method to get a BRR estimate for this variance term by keeping $\{\hat{P}_{klij}\}$ fixed in the formula in (7) throughout the replication of $\hat{\hat{N}}^{a}_{+++2}$. We write

$$\hat{V}\left[E\left[\hat{\hat{N}}_{+++2}\ \middle|\boldsymbol{T}\right]\right] = \hat{V}^{BRR}_{\boldsymbol{T}}\left[\hat{\hat{N}}_{+++2}\right] = .765 \times 10^{9} \qquad \textbf{(12)}$$

Recall the model-based and design-based components of the linearization for $\hat{\hat{N}}_{+++2}$ are uncorrelated. So substituting (11) and (12) in (9), we obtain a variance estimate for $\hat{\hat{N}}^{a}_{+++2}$ that involves both a strictly frequentist technique (BRR) and a nominally Bayesian technique (the method of Laplace). We write:

$$\hat{V}^{LBRR}\left[\hat{\hat{N}}_{+++2}\right] = \hat{V}^{L}\left[\hat{\hat{N}}_{+++2}\ \middle|\boldsymbol{T}\right] + \hat{V}^{BRR}_{\boldsymbol{T}}\left[\hat{\hat{N}}_{+++2}\right] = 2.98 \times 10^{9} \qquad \textbf{(13)}$$

The estimate in (13) obtained through our hybrid variance estimation procedure is close to that obtained strictly through BRR in (8). This suggests the large sample approximation is accurate here. This, in turn, lends some validity to the model itself. The next section attempts to further validate the model.

5. Frequentist Validation

The results in (8) suggest that the adjusted Horvitz-Thompson $\hat{N}^{a}_{+++2}$ in (6) has a substantially larger variance than the adjusted forecast estimator $\hat{\hat{N}}^{a}_{+++2}$ in (7). However, at this point it is unclear if $\hat{\hat{N}}^{a}_{+++2}$, which relies more heavily on the model, is subject to bias relative to our targeted population statistic, the number of "on layoff" at wave 2. In addition, although BRR and the method of Laplace give estimates that are close in this case, we don't know how this result could be generalized to the broader context of the full model.

These basic issues can be investigated by exploring the properties of the components of a basis for the conditional probabilities $\{P_{klij}\}$. The next section derives such a basis.

5.1 A Basis for the Transition Probabilities

We set

$$\gamma_1 = \frac{P_{1111}}{P_{1111} + P_{1112}} \;;\; \gamma_2 = \frac{P_{1121}}{P_{1121} + P_{1122}} \;;\; \gamma_3 = \frac{P_{1211}}{P_{1211} + P_{1212}}$$
$$\gamma_4 = \frac{P_{1111}}{P_{1111} + P_{1121}} \;;\; \gamma_5 = \frac{P_{2111}}{P_{2111} + P_{2121}} \tag{14}$$

Where $0 < \gamma_i < 1; \; i = 1, \ldots, 5$ Let $\boldsymbol{\gamma} = \left(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5\right)^t$ Then the components of $\boldsymbol{\gamma}$ forms a basis for the parameter space of the conditional probabilities $\left\{P_{ijkl}\right\}$, in the sense that the P_{ijkl}'s are strict algebraic functions of $\boldsymbol{\gamma}$. An algorithm for deriving $\left\{P_{ijkl}\right\}$ in terms of $\boldsymbol{\gamma}$, is presented in Thibaudeau (2003). The upshot is the basis in (19) provides us with tools to conduct model diagnoses.

5.2 Model-Based Robustness

Binder (1983) discusses weighted MLE's (WMLE) in the context of frequentist validity of model-based estimation. WMLE's are valid frequentist statistics in that they are consistent for the solution of the population likelihood equations. Pfeffermann (1993) proposes a test to assess equality in expected value between the MLE and the WMLE in our situation. The MLE and WMLE for the components of the basis $\boldsymbol{\gamma}$ are given bellow. Note the close proximity of the MLE $\hat{\gamma}_i$ to the WMLE $\tilde{\gamma}_i$, for i = 1, …, 5. The test of Pfefferman is not shown here, but is far from significant in this case.

$$\hat{\gamma}_1 = .97 \quad \tilde{\gamma}_1 = .96 \;;\; \hat{\gamma}_2 = .92 \quad \tilde{\gamma}_2 = .92 \;;\; \hat{\gamma}_3 = .36 \quad \tilde{\gamma}_3 = .35$$
$$\hat{\gamma}_4 = .79 \quad \tilde{\gamma}_4 = .80 \;;\; \hat{\gamma}_5 = .56 \quad \tilde{\gamma}_5 = .55 \tag{15}$$

What (15) suggests is the MLE and the WMLE are close in expectation. I.e. they are estimating approximately the same population statistic, namely the solution to the population likelihood equations. This gives a frequentist interpretation to the MLE $\hat{\boldsymbol{\gamma}}$. In that sense, $\hat{N}^a_{+++2}$ and $\hat{\hat{N}}^a_{+++2}$ have meaningful interpretations in the frequentist perspective.

5.3 Relative Precision of the MLE and WMLE

Given approximate equality between the MLE and the WMLE in expectation, we need a rationale for using the MLE instead of the WMLE in the first place. The rationale lies in the relative precision of the two estimators. Full replication BRR estimates using Fay's method for the variances of the MLE's of the components of the basis $\boldsymbol{\gamma}$ are given bellow, along with their counterparts for the WMLE. Based on these estimates, the MLE has a smaller variance for the first four components. The variance is essentially the same for the fifth component.

$$\hat{V}^{BRR}(\hat{\gamma}_1) = .00011 \quad \hat{V}^{BRR}(\tilde{\gamma}_1) = .00014 \;;\; \hat{V}^{BRR}(\hat{\gamma}_2) = .00070 \quad \hat{V}^{BRR}(\tilde{\gamma}_2) = .00085$$
$$\hat{V}^{BRR}(\hat{\gamma}_3) = .0059 \quad \hat{V}^{BRR}(\tilde{\gamma}_3) = .0070 \;;\; \hat{V}^{BRR}(\hat{\gamma}_4) = .00028 \quad \hat{V}^{BRR}(\tilde{\gamma}_4) = .00030 \tag{16}$$
$$\hat{V}^{BRR}(\hat{\gamma}_5) = .0035 \quad \hat{V}^{BRR}(\tilde{\gamma}_5) = .0034$$

5.4 The Method of Laplace vs. BRR

(15) and (16) lend some support to using the MLE to construct a fairly efficient estimator which is approximately unbiased for estimating our population target, the number of "on layoff" at wave 2. But we do not yet have any evidence to support using the method of Laplace, whose validity is tied to that of the model. We can use BRR as a relative gauge for the components of the basis $\boldsymbol{\gamma}$ to evaluate the method of Laplace. We have:

$$\hat{V}^{BRR}(\hat{\gamma}_1) = .00011 \quad \hat{V}^{L}(\hat{\gamma}_1) = .00010 \quad ; \quad \hat{V}^{BRR}(\hat{\gamma}_2) = .00070 \quad \hat{V}^{L}(\hat{\gamma}_2) = .00068$$
$$\hat{V}^{BRR}(\hat{\gamma}_3) = .0059 \quad \hat{V}^{L}(\hat{\gamma}_3) = .0074 \quad ; \quad \hat{V}^{BRR}(\hat{\gamma}_4) = .00028 \quad \hat{V}^{L}(\hat{\gamma}_4) = .00020 \qquad \textbf{(17)}$$
$$\hat{V}^{BRR}(\hat{\gamma}_5) = .0035 \quad \hat{V}^{L}(\hat{\gamma}_5) = .0026$$

The results in (17) suggest the method of Laplace and BRR are roughly equivalent for $\gamma_1, \gamma_2, \gamma_3$, but some distortions are evident for γ_4, γ_5. However, the accuracy of BRR is hampered by the fact there are only 23 degrees of freedom underlying the replication. Furthermore, simulations we conducted independently suggest the variance of BRR itself is large in this type of situation. The current results are not incompatible with the hypothesis of a valid model and a valid use of the method of Laplace in this situation.

6. Discussion

Our research confirms that estimators build on a structural exploitation of the model, like $\hat{\hat{N}}^{a}_{2+++}$ in (7), can lead to substantially smaller variances, relative to estimators for which the model only serves to patch-up the missing data, like $\hat{N}^{a}_{2+++}$ in (6). Also, the method of Laplace can give reasonably accurate large-sample approximations for the variances of model-based components in hybrid estimators. In turn this can lead to simplified variance estimators for the hybrids if the variance of the design-based component is also simple to estimate. The key to this approach is the validity of the model. While it is impossible to completely validate a model, we present a technique to conduct diagnostic checks that could identify some misspecifications. The expansion of a loglinear basis in (14) allows for a systematic investigation of possible relative bias between the MLE and WMLE. Evident biases would be clear indications of model misspecification.

7. References

Binder, D. A. (1983). "On the Variance of Asymptotically Normal Estimators from Complex Surveys." International Statistical Review, 51, pp279-292.

Bishop, Y. M., Fienberg , S. E., Holland, P. W. (1975). Discrete Multivariate Analysis, Theory and Practice, MIT Press.

Judkins, D. R. (1990). "Fay's Method for Variance Estimation." Journal of Official Statistics, **6,** 3.

Pfeffermann, D., Skinner, C., Humphreys, K. (1998). "The Estimation of Gross Flows in the Presence of Measurement Error Using Auxiliary Variables." JRSS A, 161, 1.

Pfeffermann, D. (1993). "The Role of Sampling Weights when Modeling Survey Data." International Statistical Review, 61, 2, pp 317-337.

Rubin, D. B. (1977). "The Design of A General and Flexible System for Handling Non-Response in Sample Surveys." Technical Report, Republished 1987.

Rubin, D. B. (1987). "Multiple Imputation for Nonresponse in Surveys." Wiley.

Thibaudeau, Y. (2003). "An Algorithm for Computing Full Rank Minimal Sufficient Statistics with Applications to Confidentiality Protection." Monographs of Official Statistics, Work Session on Statistical Data Confidentiality, Luxembourg, 7 to 9 April 2003. Part 1, pp. 45-58.
http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CR-03-004-1/EN/KS-CR-03-004-1-EN.PDF

Tierney, L., Kass, R. E., Kadane, J. B. (1989). "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions." Journal of the American Statistical Association, 84, 407.

Wolter, K. M. (1985). Introduction to Variance Estimation. Springer-Verlag Ed.