Session 5

Application of Jackknife Theory in Small Area Estimation

Jackknifing in The Fay-Herriot Model with An Example

Jiming Jiang, P. Lahiri, Shu-Mei Wan, Chien-Hua Wu

Abstract

The paper reviews empirical best linear prediction (EBLUP) and the associated jackknife MSE estimator of EBLUP. The bias of jackknife MSE estimator is of order $o(m^{-1})$, where m is the number of small areas. The jackknife works well both for normal and nonnormal Fay-Herriot models. The proposed methodology is illustrated using a real life example from the National Health and Interview Survey.

1 Introduction

Fay and Herriot (1979) put forward an empirical Bayes method to estimate per-cpita income of small-places (population less than 1000) using a Bayesian model that combines Current Population Survey data in conjunction with relevant administrative and census data. Their empirical results demonstrate that their empirical Bayes estimator performed better that both direct survey estimator and a synthetic estimator which is a direct estimator for the corresponding county. The Fay-Herriot method is a popular small-area method because of its simplicity and its demonstrated good empirical performances. It also produces design consistent estimator, a desirable property which brings a model-based estimator closer to direct estimator for large sample, irrespective of the true model.

Prasad and Rao (1990) developed a delta method for estimating mean square error (MSE)

[0] J. Jiang, Assoc. Prof., Univ. of Calif.-Davis, U.S.A; P. Lahiri, Milton Mohr Distinguished Prof., Univ. of Nebraska-Lincoln, U.S.A.; S. Wan, Assist. Prof., Lung-Hua Inst. Tech., Taiwan; C. Wu, Statistical Reviewer, Center for Drug Evaluation, Taiwan.

of empirical best linear unbiased predictor (EBLUP) of a general mixed effect in the context of a mixed *linear normal* model which covers the Fay-Herriot model. Lahiri and Rao (1995) robustified the Prasad-Rao method by allowing *nonnormal* random effects in the Fay-Herriot model. However, both the papers are vaild only for ANOVA method of estimating the model parameters. Datta and Lahiri (2000) considered the mixed linear normal model considered by Prasad and Rao (1990) but generalized the Prasad-Rao's method to include ML and REML variance component estimators. More recently, Jiang *et al.* (2001) proposed a jackknife method to estimate the MSE of empirical best predictor (EBP) for *nonnormal* and *nonlinear mixed* models and for general M-estimators of model parameters. Their MSE estimator enjoys the desirable property that the bias is of order $o(m^{-1})$.

The main purpose of this paper is to spell out the jackknife method for the Fay-Herriot model. For a very special case of the Fay-Herriot model, Lahiri (1995) noted that the jackknife MSE estimator of an EBLUP involves estimated skewness and kurtosis terms. The jackknife MSE estimtor is also asymptotically equivalent to Morris' (see Morris 1983) formula which was obtained as an approximation to the posterior variance under a uniform improper prior distribution on the model parameters. Thus, the jackknife is very similar to a Bayesian procedure, at least for this special case.

As for an illustration of our methodology, we carry out a data analysis to estimate the proportion of people who did not visit a doctor's office during the last twelve months for each state and the District of Columbia (small areas). The Fay-Herriot model cannot be applied directly to the survey estimates since one would expect its true sampling variances to be related to the corresponding true small-area proportions. Note that the Fay-Herriot model assumes

that the sampling variances to be known. Thus, we first make a suitable transformation of the direct survey estimates to stabilize their sampling variances and then assumed known design effects. We then simply apply the Fay-Herriot model on the transformed survey estimates in order to combine information from various relevant census and administrative data. The performances of the Fay-Herriot type of estimator and the associated jackknife MSE estimator seem reasonable.

2 Estimation in a non-normal Fay-Herriot Model

We assume a non-normal version of the small area model considered by Fay and Herriot (1979). According to the model, $y_i = \theta_i + e_i; \theta_i = x_i'\beta + v_i$, where e_i and v_i are all uncorrelated with $E(e_i) = E(v_i) = 0$ and $Var(e_i) = D_i,\ Var(v_i) = A\ (i = 1, \cdots, m)$. In the model, D_i's $(i = 1, \cdots, m)$ are assumed to be known. Let p be the dimension of x_i and $\phi = (\beta, A)$.

When ϕ is known the best predictor (BP) of θ_i is simply the conditional mean of θ_i given y_i and is given by $\hat{\theta}_i(y_i; \phi) = (1 - \gamma_i)x_i'\beta + \gamma_i y_i$, where $\gamma_i = A/(A + D_i)$. Note that the above BP can be also interpreted as a Bayes estimator When β is unknown but A is known, one can estimate β by the generalized least square estimator of β, given by $\hat{\beta}(A) = \left(\sum_{i=1}^m (A + D_i)^{-1} x_i x_i'\right)^{-1} \sum_{i=1}^m (A + D_i)^{-1} x_i y_i$.

An EBP [or empirical Bayes (EB)] of θ_i is then obtained by replacing β in the BP by $\hat{\beta}(A)$. Note that this is also the best linear unbiased predictor (BLUP), see Prasad and Rao (1990).

In practice A is rarely known and so it needs to be estimated from the data. Prasad and Rao (1990) used a method of moments (MOM) estimator of A, defined by $\hat{A} = max[0, \tilde{A}]$ with $\tilde{A} = (m - p)^{-1} \sum_{i=1}^m \{(y_i - x_i'\hat{\beta}_{OLS})^2 - (1 - h_i)D_i\}$, where $h_i = x_i'(\sum_{i=1}^m x_i x_i')^{-1} x_i$, $\hat{\beta}_{OLS} =$

$(\sum_{i=1}^{m} x_i x_i')^{-1} \sum_{i=1}^{m} x_i y_i$.

Researchers have used other methods of estimating A (see, *e.g.* Fay and Herriot 1979, Jiang *et al.* 2001, among others). Plugging in an estimator of A in the BLUP yields EBLUP $\hat{\theta}_i = \hat{\theta}_i(y_i; \hat{\phi}) = (1 - \hat{\gamma}_i) x_i' \hat{\beta}(\hat{A}) + \hat{\gamma}_i y_i$, where $\hat{\gamma}_i = \hat{A}/(\hat{A} + D_i)$ and

$$\hat{\beta}(\hat{A}) = (\sum_{i=1}^{m} (\hat{A} + D_i)^{-1} x_i x_i')^{-1} \sum_{i=1}^{m} (\hat{A} + D_i)^{-1} x_i y_i.$$

Note that it can be also interpreted as an EB estimator.

In order to understand if EBLUP method is effective, we now develop a method of constructing confidence intervals of γ_i based on the point estimates $\hat{\gamma}_i$ $(i = 1, \cdots, m)$. Applying Taylor series method, we obtain an estimator of $Var(\hat{\gamma}_i)$ as $v_J(\hat{\gamma}_i) = \frac{D_i^2}{(\hat{A}+D_i)^4} v_J(\hat{A})$, where $v_J(\hat{A}) = \frac{m-1}{m} \sum_{u=1}^{m} (\hat{A}_{-u} - \hat{A})^2$ is a jackknife estimator of $Var(\hat{A})$. Here $\hat{A}_{-u}$ is calculated exactly in the same way as A except that the data for the uth small area is deleted $(u = 1, \cdots, m)$ We can construct the confidence intervals of γ_i as $\{\hat{\gamma}_i - 2\sqrt{v_J(\hat{\gamma}_i)}, \hat{\gamma}_i + 2\sqrt{v_J(\hat{\gamma}_i)}\}$, $i = 1, \cdots, m$. If EBLUP is effective, the confidence intervals of γ_i for most of the states will not contain 1 or 0.

Next, we discuss the important assumption of known sampling variances D_i's. In Fay and Herriot (1979), the following justification was given. They assumed that the coefficient of variation for the i th small area direct estimate is approximately $3/\sqrt{N_i}$, where N_i denotes 20 percent sample count. This approximation was made based on an empirical study which found that above approximation works well for eight states. A log transformation was then taken to stabilize the variance and the Fay-Herriot model was applied on the transformed direct estimates with $D_i = 9/N_i$.

In view of the above discussion, we are often encountered with the problem of estimation

of $h(\theta_i)$, a function, possibly nonlinear, in θ_i. A simple estimator (*e.g.,* Fay and Herriot 1979) is $h^{-1}(\hat{\theta}_i)$, where $h^{-1}(.)$ is the inverse transformation of $h(.)$. This is not an EBP but should work fine as long as sample size for the small areas are not very small. We, however, note that it is possible to come up to the BP of $h(\theta_i)$ and hence EBP (see, Lahiri 1999).

3 Jackknifing in the Fay-Herriot Model

In this section, we spell out the jackknife MSE estimator of the Fay-Herriot type estimator based on Jiang *et al* (2001). The MSE of $\hat{\theta}_i$ is defined as $MSE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where E is with respect to the Fay-Herriot mixed model. Note that the MSE of the BP is given by $g_{1i}(A) = A(1-\gamma_i)$. One can then naively propose a MSE estimator of EBLUP as $g_{1i}(\hat{A})$. The problem with this naive estimator is that it does not incorporate the extra variabilities due to the estimation of ϕ and so underestimates the true MSE. Several researchers addressed this important issue and came up with improved MSE estimators which account for these extra variabilities (see, *e.g.,* Prasad and Rao 1990, Lahiri and Rao 1995, Datta and Lahiri 2000, among others). But they are all valid for mixed linear normal model.

Recently, Jiang *et al.* (2001) proposed a jackknife method which takes into account uncertainties due to the estimation of ϕ. This method is valid for nonnormal and nonlinear mixed models and for general M-estimators of model parameters. For the Fay-Herriot model, the jackknife MSE estimator of EBLUP is given by

$$mse_J(\hat{\theta}_i) = g_{1i}(\hat{A}) - \frac{m-1}{m}\sum_{u=1}^{m}[g_{1i}(\hat{A}_{-u}) - g_{1i}(\hat{A})] + \frac{m-1}{m}\sum_{u=1}^{m}(\hat{\theta}_{i,-u} - \hat{\theta}_i)^2,$$

where $\hat{A}_{-u}(\hat{\beta}_{-u})$ is the estimator of A (β) after deleting the u^{th} small-area data,

$$\hat{\theta}_{i,-u} \;=\; \hat{\gamma}_{i,-u}y_i + (1-\hat{\gamma}_{i,-u})x_i'\hat{\beta}_{-u},$$

$$\begin{aligned}\hat{\gamma}_{i,-u} &= \frac{\hat{A}_{-u}}{\hat{A}_{-u}+D_i},\\ g_{1i}(\hat{A}_{-u}) &= \hat{A}_{-u}(1-\hat{\gamma}_{i,-u}).\end{aligned}$$

Lahiri (1995) examined the above jackknife MSE estimator for a very special case of the Fay-Herriot model with $D_i = D$ and $x_i'\beta = \mu$ $(i = 1, \cdots, m)$. He showed that

$$\begin{aligned}\widehat{MSE}(\hat{\theta}_i) &= g_1(\hat{A}) + g_2(\hat{A}) + \frac{D^2}{m(\hat{A}+D)}(b_2-1)\\ &\quad + \frac{D^2}{m(\hat{A}+D)^2}(b_2-1)(y_i-\bar{y})^2 - \frac{2D^2}{m(\hat{A}+D)^{3/2}}\sqrt{b_1}(y_i-\bar{y}),\end{aligned}$$

where $b_1 = m_3^2/(\hat{A}+D)^3$, $b_2 = m_4/(\hat{A}+D)^2$ and $g_2(A) = \frac{D_i^2}{(A+D_i)^2}x_i'(\sum_{u=1}^m \frac{x_u x_u'}{A+D_u})^{-1}x_i$. Thus, unlike the normality-based MSE estimators, the jackknife MSE estimator involves estimated skewness and kurtosis terms. Lahiri (1995) also compared jackknife with the two normality-based MSE estimators of the following EBP of θ_i. $\hat{\theta}_i = \bar{y} + (1-\hat{B}_1)(y_i-\bar{y})$, where $\bar{y} = m^{-1}\sum_{i=1}^m y_i$ and $\hat{B} = \frac{D(m-3)}{\sum_{i=1}^m (y_i-\bar{y})^2}$.

We present three formulae below for comparison:

Morris (1983):

$$(1-\hat{B}_1) + \frac{D\hat{B}_1}{m} + \frac{2\hat{B}_1{}^2}{m-3}(y_i-\bar{y})^2,$$

Prasad and Rao (1990):

$$(1-\hat{B}_1)D + \frac{D\hat{B}_1}{m} + \frac{2D\hat{B}_1}{m},$$

The jackknife Formula:

$$(1-\hat{B}_1)D + \frac{D\hat{B}_1}{m} + \frac{2\hat{B}_1^2}{m}(y_i-\bar{y})^2.$$

The jackknife estimator is equivalent to the Morris' formula which was obtained as an approximation to the posterior variance formula under uniform improper prior on the model

parameters: μ and A. Also, the bias of our jackknife MSE estimator is of the order $o(m^{-1})$. Thus, our jackknife MSE estimator enjoys both good frequentist and Bayesian properties. It is interesting to note that the Prasad-Rao MSE estimator, unlike Morris' and ours, is the same for all the areas in this balanced case.

The above results are for the transformed scale. We need to provide results in the original scale. We approximate the MSE of $h^{-1}(\hat{\theta}_i)$ by $mse[h^{-1}(\hat{\theta}_i)] = [h^{-1\prime}(\hat{\theta}_i)]mse(\hat{\theta}_i)$, where $h^{-1\prime}(x)$ denotes the derivative of $h^{-1}(x)$ with respect to x and $mse(\hat{\theta}_i)$ is an estimate of MSE obtained by jackknife method as described above.

4 Data Analysis

In this section, we demonstrate our methodology to estimate the proportion of individuals who did not visit doctor's office during the last twelve months for all the fifty states and the District of Columbia (small areas) using the National Health Interview Survey (NHIS) data in conjunction with relevant administrative and census data. Earlier Malec *et al.* (1997) proposed a hierarchical Bayes method to address the same estimation problem. Unlike our modeling, they used an individual level model and their method does not produce design consistent small area estimators. Also, they did not use auxiliary data at the small area level. Our method can be viewed as a first step in getting a simple minded design consistent small area estimators. It has also a huge computational advantage over their procedure.

Let n_i be the sample size for the ith state and w_{ij} be the sampling weight for the jth individual in the ith state ($i = 1, \cdots, m = 51; j = 1, \cdots, n_i$). For the jth individual in the ith state, we observe a binary response z_{ij} which takes on the value 1 if the individual did not

visit a doctor's office during the last 12 months and 0 otherwise ($i = 1, \cdots, m; j = 1, \cdots, n_i$). Then, $z_i = \sum_{j=1}^{n_i} w_{ij} z_{ij} / \sum_{j=1}^{n_i} w_{ij}$ is the direct survey estimate of π_i, the true proportion of individuals who did not visit a doctor's office during the last 12 months for the ith state ($i = 1, \cdots, m$). Using SUDDAN, sample survey software, the NCHS has provided data on z_i and its sampling variance V_i for $i = 1, \cdots, m$.

Note that $E(z_i|\pi_i) \approx \pi_i$, and

$$V(z_i|\pi_i) = D^{\star}_{i0} \cdot \frac{\pi_i(1-\pi_i)}{n_i}$$

where

$$D^{\star}_{i0} = \frac{V_i^*}{\frac{\pi_i(1-\pi_i)}{n_i}} \approx \frac{V_i}{\frac{z_i(1-z_i)}{n_i}} D_{i0}.$$

The factor $D^{\star}_{i0}$ is known as a design effect and adjusts the simple random sampling formula by incorporating effects due to clustering and unequal probability selections. Consider the transformation: $y_i = \sin^{-1}\sqrt{z_i}$. Thus, for this example, $h(.) = \sin^{-1}(.)$. By Taylor series argument, we have

$$E(y_i|\theta_i) \approx \sin^{-1}\sqrt{\pi_i} = \theta_i,$$

and

$$\begin{aligned} V(y_i|\theta_i) &= V[\sin^{-1}\sqrt{z_i}|\theta_i] \\ &\approx V[\sin^{-1}\sqrt{\pi_i} + (z_i - \pi_i) \cdot \frac{1}{2\sqrt{\pi_i(1-\pi_i)}}] \\ &= \frac{1}{4\pi_i(1-\pi_i)} \cdot D^{\star}_{i0} \cdot \frac{\pi_i(1-\pi_i)}{n_i} \\ &= \frac{D^{\star}_{i0}}{4n_i}. \end{aligned}$$

We will assume that $D_i = \frac{D_{i0}}{4n_i}$ is the estimated sampling variance of y_i $(i = 1, \cdots, m)$.

In addition to z_i and V_i provided to us by the NCHS, we collected data on 1990 urban population (X_1), 1995 Bachelor's degree completion for 25+ population (X_2), 1995 high school completion for the 25+ population (X_3), 1995 health insurance coverage (X_4), and 1990 physician population (X_5) for each of the 50 states and the District of Columbia. The covariate information was obtained from the Census Bureau web site.

We first consider the issue of covariate selection. For this purpose, we considered the largest 15 states (in terms of sample size) and used SAS to produce Tables 1 and 2. For these states sampling variabilities are very low and so the usual SAS procedures are justified. Note that correlations between Y and each of the three covariates X_2, X_3, and X_4 are significant (at 0.1 level). While X_2 is significantly correlated with X_3, it is not significantly correlated with X_4. Likewise, X_4 is significantly correlated with X_3, but not with X_2. Thus, keeping the aspect of multicollinearity in mind, Table 1 suggests to consider X_2 and X_4 in the model. The selection of these two covariates is confirmed by Table 2. Both R^2 and adjusted R^2 are the highest when we include X_2 and X_4 in the model. We would also select these two covariates when we apply the C_p criterion.

The estimates of γ ranges between .09 (South Dakota) and .95 (California), depending on the sampling variability of the corresponding state NCHS estimate. The γ values are low for small states (indicating that the NCHS estimates are highly unreliable) and large for large states (indicating the NCHS estimates are reliable). None of the confidence intervals [(L.L.,U.L)] includes 0 or 1 suggesting the use of EBLUP. See Table 3 and Figure 3.

The estimator of π_i is then given by $\hat{\pi}_i = \sin^2(\hat{\theta}_i)$. Thus, here $h^{-1}(.) = \sin^2(.)$. The MSE

of $\hat{\pi}_i$ is given by $4\hat{\pi}_i(1-\hat{\pi}_i)mse(\hat{\theta}_i)$. A synthetic estimator of π_i is given by $\hat{\pi}_i^{syn} = \sin^2(x_i'\hat{\beta})$.

Table 4 reports the NCHS estimates (z), proposed composite estimates ($\hat{\pi}$), and synthetic estimates ($\hat{\pi}^{syn}$). For large states (e.g., California , Texas, etc.), our proposed composite estimates are similar to the NCHS estimates. Figure 1 plots these estimates.

Finally, Tables 5 provides standard errors of the NCHS estimates ($se(z) = \sqrt{V_i}$), the jackknife MSE estimates of our proposed composite estimates ($se(\hat{\pi}) = \sqrt{mse_J(\hat{\pi})}$), and percent improvement defined by $PCTIMP = 100 \times \frac{se(z)-se(\hat{\pi})}{se(z)}$. A corresponding plot of $PCTIMP$ is given in Figure 3. For small states (e.g., South Dakota, Vermont, etc.), improvement is quite substantial.

Acknowledgements

The research was supported, in part, by the National Science Foundation Grants and a contract from the National Center for Health Statistics.

REFERENCES

Chattopadhyay, M., Lahiri, P., Larsen, M., and Reimnitz, J. (1999), Composite estimation of drug prevalences for sub-state areas, *Survey Methodology*, 25, 81-86.

Datta, G.S. and Lahiri, P. (2000), A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems *Statist. Sinica*.

Jiang, J., Lahiri, P., and Wan, S. (2001), A unified jackknife theory, under revision for *Ann. Statist.*

Fay, R. E., and Herriot, R. A. (1979), Estimates of income for small places: an application

of James-Stein procedures to census data, *J. Amer. Statist. Assoc.* 74, 269-277.

Lahiri, P. (1995), A jackknife measure of uncertainty of linear empirical Bayes estimators, unpublished manuscript.

Lahiri, P. (1999), Discussion on *Current Trends in Sample Survey* by J.N.K. Rao, *Sankhya*.

Lahiri, P., and Rao, J.N.K. (1995), Robust estimation of mean squared error of small area estimators, *J. Amer. Statist. Asso.*, 90, 758-766.

Malec, D., Sedransk, J., Moriarity, C.L., and LeClere, F.B. (1997), Small area inference for binary variables in the National Health Interview Survey, *J. Amer. Statist. Asso.* 92, 815-826.

Morris, C. (1983), Parametric empirical Bayes inference: theory and application, *J. Amer. Statist. Asso.,* 78, 47-59.

Prasad, N.G.N., and Rao, J.N.K. (1990), The estimation of mean squared errors of small area estimators, *J. Amer. Statist. Asso.*, 85, 163-171.

Table 1: Pearson Correlation Coefficients

	Y	X_1	X_2	X_3	X_4	X_5
Y	1.00000	0.10733	0.59754	0.55709	0.51929	0.26759
	0.0	0.7034	0.0187	0.0310	0.0473	0.3349
X_1	0.10733	1.00000	0.56726	0.44259	-0.25401	-0.64715
	0.7034	0.0	0.0274	0.0985	0.3610	0.0091
X_2	0.59754	0.56726	1.00000	0.45595	0.02605	0.05537
	0.0187	0.0274	0.0	0.0876	0.9266	0.8446
X_3	0.55709	0.44259	0.45595	1.00000	0.62160	-0.09249
	0.0310	0.0985	0.0876	0.0	0.0134	0.7430
X_4	0.51929	-0.25401	0.02605	0.62160	1.00000	0.26579
	0.0473	0.3610	0.9266	0.0134	0.0	0.3383
X_5	0.26759	-0.64715	0.05537	-0.09249	0.26579	1.00000
	0.3349	0.0091	0.8446	0.7430	0.3383	0.0

Table 2: Values of C_p statistic, R^2, and Adjusted R^2 for different possible models

Model	Variables	C_p	R^2	Adjusted R^2
1	X_2	7.23672	0.35705587	0.3076
2	X_3	8.56150	0.31035017	0.2573
3	X_4	9.71567	0.26965933	0.2135
4	X_2, X_3	6.33543	0.45934214	0.3692
5	X_2, X_4	2.03477	0.61096386	0.5461
6	X_3, X_4	9.17802	0.35912540	0.2523
7	X_2, X_3, X_4	4.00000	0.61218982	0.5064

Table 3: Direct survey estimates (y), synthetic estimates ($x'\hat{\beta}$), EBLUP's ($\hat{\theta}$), and confidence Intervals of γ's

	STATE	y	$x'\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	L.L.	U.L.
1	Alabama	1.08078	1.04874	0.85047	1.07599	0.72554	0.97539
2	Alaska	0.99007	1.08821	0.31453	1.05734	0.10274	0.52632
3	Arizona	1.05567	1.02178	0.73862	1.04681	0.54898	0.92827
4	Arkansas	1.05055	1.01292	0.72774	1.04031	0.53310	0.92237
5	California	1.04064	1.04298	0.95314	1.04075	0.90926	0.99701
6	Colorado	1.09146	1.11180	0.63604	1.09886	0.40864	0.86344
7	Connecticut	1.16463	1.13980	0.67252	1.15649	0.45618	0.88887
8	Delaware	1.12824	1.06208	0.29641	1.08169	0.09154	0.50127
9	D.C.	1.19544	1.12036	0.28924	1.14207	0.08729	0.49118
10	Florida	1.00245	1.04526	0.82933	1.00976	0.69029	0.96837
11	Georgia	1.06529	1.04995	0.70721	1.06080	0.50380	0.91061
12	Hawaii	1.10652	1.09525	0.45527	1.10038	0.21166	0.69889
13	Idaho	1.03265	1.06746	0.37944	1.05425	0.14814	0.61074
14	Illinois	1.03504	1.09350	0.84473	1.04412	0.71589	0.97357
15	Indiana	1.00212	1.05179	0.61762	1.02111	0.38563	0.84961
16	Iowa	1.13109	1.07058	0.47988	1.09962	0.23469	0.72506
17	Kansas	1.07790	1.09118	0.62394	1.08289	0.39345	0.85443
18	Kentucky	1.04731	1.05219	0.72660	1.04865	0.53145	0.92174
19	Louisiana	0.99794	1.02531	0.69639	1.00625	0.48870	0.90409
20	Maine	1.06623	1.06694	0.69604	1.06645	0.48821	0.90387
21	Maryland	1.11723	1.07953	0.76824	1.10849	0.59334	0.94314
22	Massachusetts	1.15534	1.12785	0.83892	1.15091	0.70618	0.97166
23	Michigan	1.06529	1.08287	0.80384	1.06874	0.64894	0.95873
24	Minnesota	1.08561	1.11693	0.85396	1.09019	0.73145	0.97647
25	Mississippi	0.99608	1.01842	0.71197	1.00251	0.51053	0.91342

Table 3 continued

	STATE	y	$x'\hat{\beta}$	γ	$\hat{\theta}$	L.L.	U.L.
26	Missouri	1.05847	1.06318	0.80648	1.05938	0.65316	0.95979
27	Montana	1.05871	1.07378	0.52071	1.06593	0.27555	0.76587
28	Nebraska	1.07682	1.10131	0.50497	1.08894	0.25941	0.75053
29	Nevada	1.03356	1.02295	0.31109	1.02625	0.10056	0.52161
30	New Hampshire	1.13083	1.10371	0.52596	1.11797	0.28104	0.77088
31	New Jersey	1.07383	1.09149	0.80306	1.07731	0.64769	0.95842
32	New Mexico	0.92677	1.00499	0.45629	0.96930	0.21258	0.69999
33	New York	1.09796	1.07945	0.92424	1.09656	0.85545	0.99302
34	North Carolina	1.03038	1.05896	0.79596	1.03621	0.63642	0.95550
35	North Dakota	1.08139	1.08619	0.22328	1.08512	0.05292	0.39364
36	Ohio	1.10328	1.06747	0.84758	1.09783	0.72067	0.97448
37	Oklahoma	1.05532	1.02779	0.76275	1.04879	0.58499	0.94052
38	Oregon	1.04041	1.07866	0.62788	1.05464	0.39836	0.85740
39	Pennsylvania	1.10080	1.08103	0.82972	1.09744	0.69093	0.96851
40	Rhode Island	1.19764	1.09811	0.42828	1.14074	0.18775	0.66881
41	South Carolina	1.05392	1.04741	0.71028	1.05203	0.50814	0.91243
42	South Dakota	1.03038	1.07693	0.09482	1.07252	0.01051	0.17913
43	Tennessee	1.09024	1.04420	0.79321	1.08071	0.63208	0.95434
44	Texas	1.00598	1.01310	0.94163	1.00640	0.88764	0.99562
45	Utah	1.03243	1.08697	0.47525	1.06105	0.23027	0.72023
46	Vermont	1.06659	1.10686	0.19413	1.09904	0.04045	0.34781
47	Virginia	1.09292	1.08670	0.81669	1.09178	0.66964	0.96375
48	Washington	1.08513	1.09427	0.78455	1.08710	0.61850	0.95059
49	West Virginia	1.03618	1.01954	0.32898	1.02501	0.11213	0.54583
50	Wisconsin	1.08380	1.09456	0.81777	1.08576	0.67138	0.96416
51	Wyoming	1.08030	1.05342	0.52217	1.06746	0.27707	0.76727

Table 4: Direct estimates (z), composite estimates ($\hat{\pi}$), and synthetic estimates ($\hat{\pi}^{syn}$)

	State	z	$\hat{\pi}$	$\hat{\pi}^{syn}$
1	Alabama	0.7785	0.77451	0.75134
2	Alaska	0.6990	0.75873	0.78464
3	Arizona	0.7573	0.74967	0.72768
4	Arkansas	0.7529	0.74401	0.71975
5	California	0.7443	0.74440	0.74634
6	Colorado	0.7873	0.79333	0.80371
7	Connecticut	0.8439	0.83795	0.82546
8	Delaware	0.8166	0.77925	0.76277
9	District of Columbia	0.8656	0.82719	0.81046
10	Florida	0.7103	0.71691	0.74832
11	Georgia	0.7655	0.76168	0.75238
12	Hawaii	0.7995	0.79456	0.79040
13	Idaho	0.7373	0.75609	0.76734
14	Illinois	0.7394	0.74733	0.78897
15	Indiana	0.7100	0.72708	0.75397
16	Iowa	0.8188	0.79394	0.76997
17	Kansas	0.7761	0.78025	0.78707
18	Kentucky	0.7501	0.75125	0.75431
19	Louisiana	0.7062	0.71374	0.73082
20	Maine	0.7663	0.76648	0.76690
21	Maryland	0.8080	0.80107	0.77746
22	Massachusetts	0.8371	0.83382	0.81630
23	Michigan	0.7655	0.76842	0.78023
24	Minnesota	0.7825	0.78626	0.80777
25	Mississippi	0.7045	0.71036	0.72468

Table 4 continued

	State	z	$\hat{\pi}$	$\hat{\pi}^{syn}$
26	Missouri	0.7597	0.76048	0.76371
27	Montana	0.7599	0.76604	0.77265
28	Nebraska	0.7752	0.78524	0.79531
29	Nevada	0.7381	0.73165	0.72872
30	New Hampshire	0.8186	0.80859	0.79724
31	New Jersey	0.7727	0.77561	0.78733
32	New Mexico	0.6395	0.67978	0.71260
33	New York	0.7926	0.79146	0.77739
34	North Carolina	0.7353	0.74043	0.76012
35	North Dakota	0.7790	0.78209	0.78297
36	Ohio	0.7969	0.79249	0.76735
37	Oklahoma	0.7570	0.75138	0.73301
38	Oregon	0.7441	0.75642	0.77673
39	Pennsylvania	0.7949	0.79217	0.77871
40	Rhode Island	0.8671	0.82618	0.79272
41	South Carolina	0.7558	0.75418	0.75018
42	South Dakota	0.7353	0.77160	0.77529
43	Tennessee	0.7863	0.77844	0.74740
44	Texas	0.7135	0.71388	0.71991
45	Utah	0.7371	0.76190	0.78362
46	Vermont	0.7666	0.79348	0.79977
47	Virginia	0.7885	0.78757	0.78339
48	Washington	0.7821	0.78372	0.78960
49	West Virginia	0.7404	0.73055	0.72568
50	Wisconsin	0.7810	0.78262	0.78983
51	Wyoming	0.7781	0.76733	0.75537

Table 5: Design effects (D_0), standard errors of direct estimates [$se(z)$], and jackknife standard errors of composite estimates [$se(\hat{\pi})$], and the percent improvement ($PCTIMP$)

STATE	D_0	$se(z)$	$se(\hat{\pi})$	$CV(z)$	$CV(\hat{\pi})$	$PCTIMP$
1	0.94237	0.0099	0.009678	0.01272	0.012495	2.2442
2	1.00743	0.0385	0.027225	0.05508	0.035889	29.2862
3	1.77080	0.0145	0.013842	0.01915	0.018464	4.5357
4	0.97357	0.0150	0.014295	0.01992	0.019212	4.7026
5	2.33951	0.0055	0.005454	0.00739	0.007327	0.8372
6	2.28631	0.0176	0.015935	0.02235	0.020087	9.4589
7	2.15494	0.0144	0.013541	0.01706	0.016159	5.9646
8	1.89534	0.0339	0.027026	0.04151	0.034678	20.2782
9	1.86680	0.0304	0.024869	0.03512	0.030061	18.1926
10	3.67215	0.0117	0.011234	0.01647	0.015671	3.9842
11	3.32450	0.0155	0.014593	0.02025	0.019159	5.8500
12	2.07315	0.0249	0.021185	0.03114	0.026662	14.9180
13	2.14117	0.0320	0.025036	0.04340	0.033115	21.7635
14	2.47474	0.0107	0.010277	0.01447	0.013752	3.9540
15	4.29302	0.0203	0.018116	0.02859	0.024918	10.7605
16	2.97118	0.0228	0.020464	0.02785	0.025772	10.2464
17	1.82169	0.0184	0.016652	0.02371	0.021342	9.5004
18	1.76132	0.0151	0.014196	0.02013	0.018897	5.9837
19	2.16050	0.0171	0.015837	0.02421	0.022190	7.3858
20	0.61690	0.0159	0.014833	0.02075	0.019352	6.7100
21	1.63347	0.0123	0.011871	0.01522	0.014819	3.4840
22	1.38477	0.0092	0.008985	0.01099	0.010775	2.3375
23	2.77288	0.0119	0.011386	0.01555	0.014818	4.3202
24	0.93043	0.0097	0.009371	0.01240	0.011918	3.3941
25	1.11029	0.0165	0.015381	0.02342	0.021653	6.7832

Table 5 continued

STATE	D_0	$se(z)$	$se(\hat{\pi})$	$CV(z)$	$CV(\hat{\pi})$	$PCTIMP$
26	1.61813	0.0119	0.011430	0.01566	0.015030	3.9498
27	1.06524	0.0233	0.020155	0.03066	0.026312	13.4962
28	1.85705	0.0235	0.020023	0.03031	0.025500	14.7958
29	3.22858	0.0372	0.028319	0.05040	0.038705	23.8739
30	1.13044	0.0208	0.018578	0.02541	0.022975	10.6826
31	2.57178	0.0118	0.011286	0.01527	0.014551	4.3583
32	4.22565	0.0298	0.024433	0.04660	0.035949	18.0113
33	1.91692	0.0066	0.006523	0.00833	0.008242	1.1667
34	2.07917	0.0127	0.012104	0.01727	0.016349	4.6896
35	2.41776	0.0440	0.029636	0.05648	0.037893	32.6465
36	2.18351	0.0097	0.009491	0.01217	0.011976	2.1512
37	1.41874	0.0136	0.013033	0.01797	0.017345	4.1681
38	2.39483	0.0191	0.017143	0.02567	0.022665	10.2484
39	2.74191	0.0104	0.010104	0.01308	0.012755	2.8445
40	1.56647	0.0223	0.020640	0.02572	0.024979	7.4418
41	1.58622	0.0156	0.014652	0.02064	0.019428	6.0756
42	8.20855	0.0775	0.035798	0.10540	0.046397	53.8087
43	1.65433	0.0119	0.011554	0.01513	0.014842	2.9060
44	1.84504	0.0064	0.006331	0.00897	0.008869	1.0754
45	2.30226	0.0263	0.021702	0.03568	0.028487	17.4841
46	3.52922	0.0490	0.030204	0.06392	0.038067	38.3593
47	1.68620	0.0110	0.010620	0.01395	0.013484	3.4547
48	1.64145	0.0123	0.011732	0.01573	0.014969	4.6200
49	3.36938	0.0356	0.027687	0.04808	0.037898	22.2287
50	1.37013	0.0111	0.010674	0.01421	0.013639	3.8400
51	0.36386	0.0226	0.020070	0.02905	0.026154	11.1955

Table 6: States arranged in increasing order of V_i

Rank of V_i	State	State ID	V_i
1	California	5	.0000303
2	Texas	44	.0000410
3	New York	33	.0000436
4	Massachusetts	22	.0000846
5	Minnesota	24	.0000941
6	Ohio	36	.0000941
7	Alabama	1	.0000980
8	Pennsylvania	39	.0001082
9	Illinois	14	.0001145
10	Virginia	47	.0001210
11	Wisconsin	50	.0001232
12	Florida	10	.0001369
13	New Jersey	31	.0001392
14	Michigan	23	.0001416
15	Missouri	26	.0001416
16	Tennessee	43	.0001416
17	Maryland	21	.0001513
18	Washington	48	.0001513
19	North Carolina	34	.0001613
20	Oklahoma	37	.0001850
21	Connecticut	7	.0002074
22	Arizona	3	.0002103
23	Arkansas	4	.0002250
24	Kentucky	18	.0002280
25	Georgia	11	.0002403

Table 6 continued

Rank of V_i	State	State ID	V_i
26	South Carolina	41	.0002434
27	Maine	20	.0002528
28	Mississippi	25	.0002723
29	Louisiana	19	.0002924
30	Colorado	6	.0003098
31	Kansas	17	.0003386
32	Oregon	38	.0003648
33	Indiana	15	.0004121
34	New Hampshire	30	.0004326
35	Rhode Island	40	.0004973
36	Wyoming	51	.0005108
37	Iowa	16	.0005198
38	Montana	27	.0005429
39	Nebraska	28	.0005523
40	Hawaii	12	.0006200
41	Utah	45	.0006917
42	New Mexico	32	.0008880
43	District of Columbia	9	.0009242
44	Idaho	13	.0010240
45	Delaware	8	.0011492
46	West Virginia	49	.0012674
47	Nevada	29	.0013838
48	Alaska	2	.0014823
49	North Dakota	35	.0019360
50	Vermont	46	.0024010
51	South Dakota	42	.0060063

Figure 1: Estimates [direct (D), synthetic (S), and EBLUP (C)] Plotted against States Arranged in Increasing Order of V_i (see Table 6 for identifying the states)

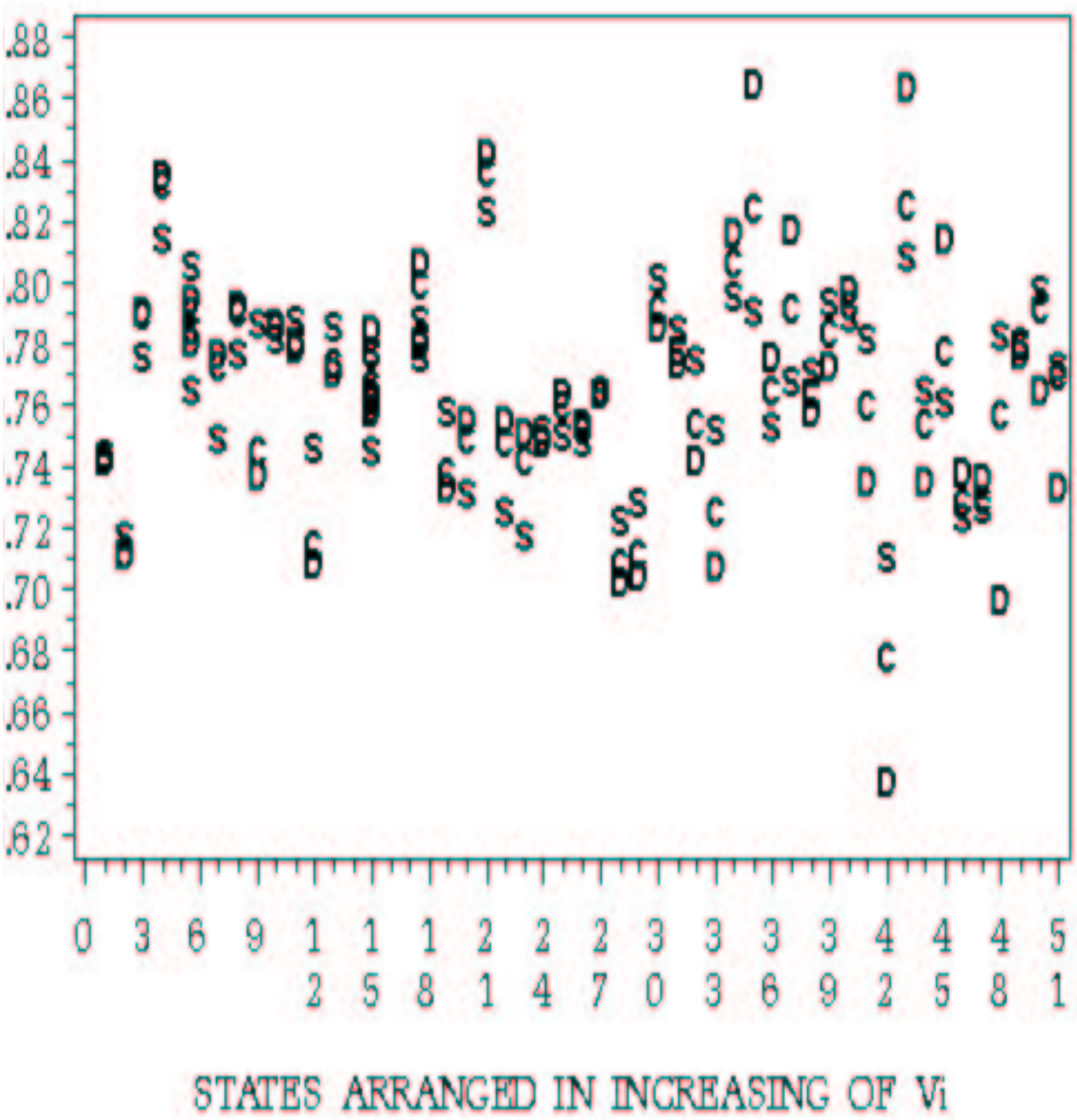


Figure 2: Percent Improvement Plotted against States Arranged in Increasing Order of V_i (see Table 6 for identifying the states)

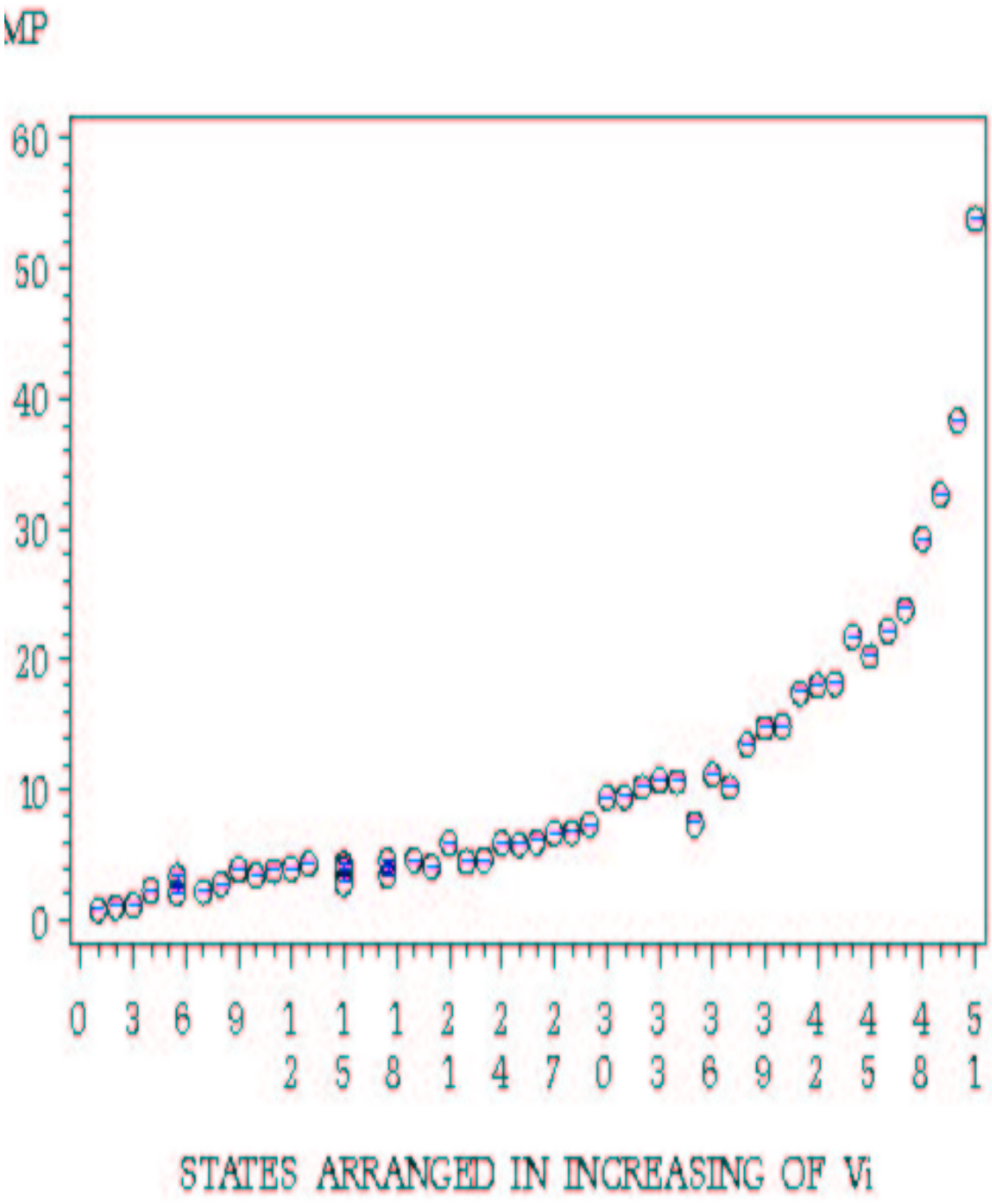
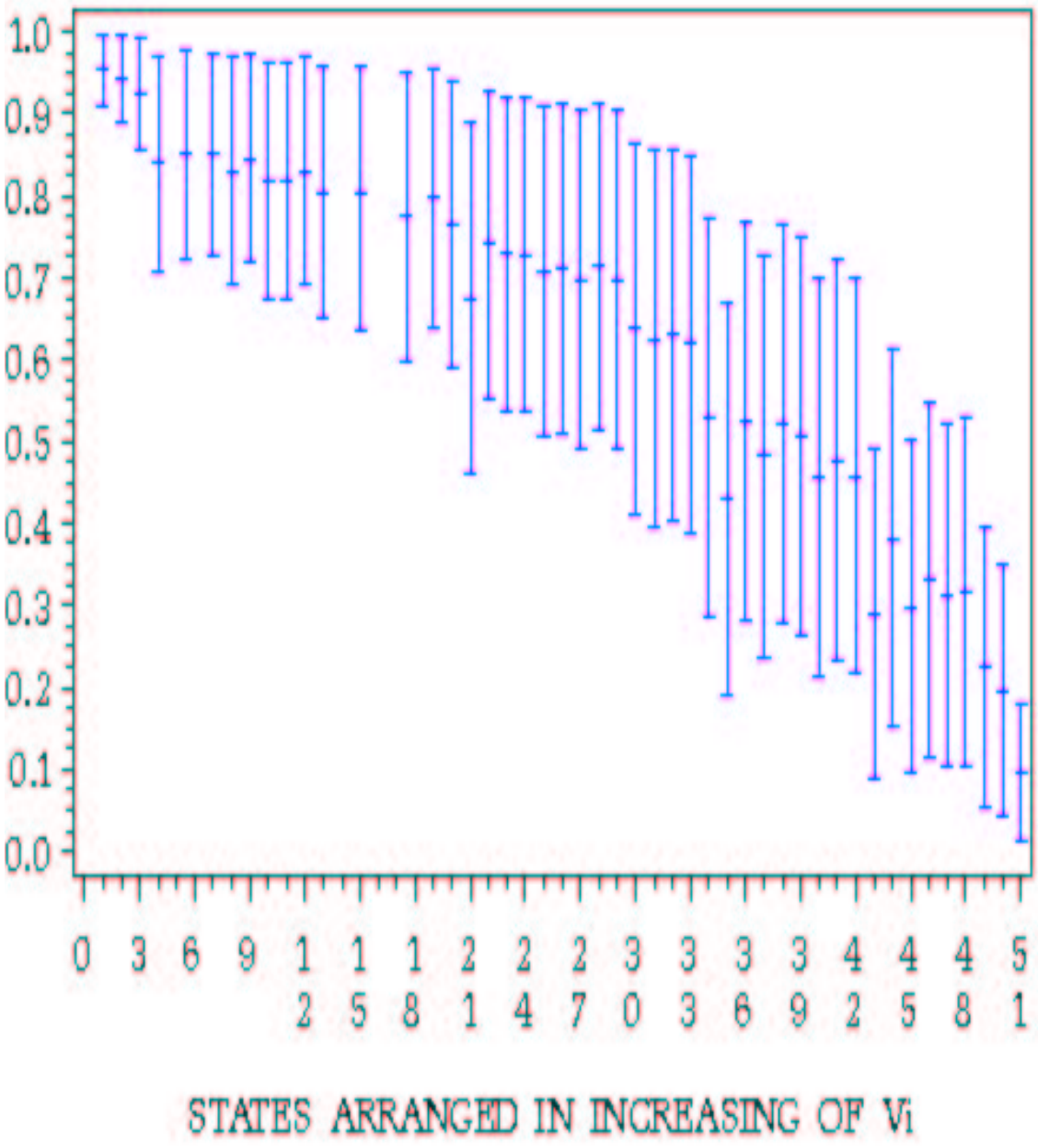


Figure 3: Confidence Interval for γ Plotted against States Arranged in Increasing Order of V_i (see Table 6 for identifying the states)



Jackknifing in the Fay-Herriott Model with An Example

by Jiming Jiang, Partha Lahiri, Shu-Mei Wan, and Chien-Hua Wu

Discussion (corrected version, 10/19/2001)

by William R. Bell, U.S. Bureau of the Census

The paper by Jiang, Lahiri, Wan, and Wu (hereafter JLWW) considers use of the jackknife to estimate the mean squared error (MSE) of small area estimates from Fay-Herriott (1979) models. The paper notes that Jiang *et al.* (2001) discuss use of the jackknife more generally for estimating MSE with nonlinear and nonnormal small area models. As the present paper restricts consideration to the linear model case, my remarks will focus only on this case. It should be kept in mind, however, that ignoring the generality of the jackknife may be ignoring one of its prime advantages.

The model of Fay and Herriott (1979) for small area estimation can be written

$$\begin{aligned} y_i &= \theta_i + e_i \qquad i = 1, \ldots, m & (1) \\ &= (x_i'\beta + v_i) + e_i & (2) \end{aligned}$$

JLWW give detailed assumptions underlying this model. Here I simply repeat that the sampling variances $D_i = \mathrm{Var}(e_i)$ of the direct survey estimates y_i are assumed known (actually meaning they are estimated using survey microdata), so that the unknown parameters of the model given by (1) and (2) are the regression parameters β and the model error variance $A = \mathrm{Var}(v_i)$. To apply this model from a frequentist perspective one first estimates the model parameters β and A using the direct estimates y_i, and then applies standard empirical Bayes prediction formulas to produce point estimates of the θ_i. A Bayesian approach can also be used (Berger 1985, Bell 1999).

Assuming the model given by (1) and (2) is true (more on this later), the error in the estimates of the θ_i can be broken into three terms:

$$\begin{aligned} \text{error} &= \text{error when all parameters are known} \\ &\quad + \text{contribution to error from estimating } \beta \\ &\quad + \text{contribution to error from estimating } A \end{aligned} \qquad (3)$$

The mean square of this error for area i is, under suitable assumptions,

$$\begin{aligned} MSE_i &= g_{1i}(A) + g_{2i}(A) + g_{3i}(A) \\ &= A(1-\gamma_i) + (1-\gamma_i)^2 x_i' \mathrm{Var}\left(\widehat{\beta}\right) x_i + g_{3i}(A) \qquad (4) \end{aligned}$$

where

$$\gamma_i = A/(A + D_i)$$

Since A is unknown to estimate MSE we plug an estimate of A into (4). Results for the term $g_{3i}(A)$ are discussed shortly. The "suitable assumptions" referred to above include normality, which is relevant in regard to asymptotic orthogonality of the second and third terms in (3). My focus, however, will be on comparing the form of (4) with the jackknife estimate of MSE suggested by JLWW, and on examining results from the jackknife for a particular empirical example. For simplicity of notation I will henceforth drop the subscript i that indexes the small areas. It should be understood that all subsequent expressions implicitly depend on i, e.g., through x_i and D_i.

Two problems arise in applying (4):

- $g_1(\widehat{A})$ is biased. In fact, even when $\widehat{A}$ is approximately unbiased, $E[g_1(\widehat{A})] \approx g_1(A) - g_3(A)$.
- There is no exact formula for $g_3(A)$.

Several approaches have previously been suggested to deal with the second problem: ignore $g_3(A)$ (naive approach); estimate $g_3(A)$ using an asymptotic expression (Prasad and Rao 1990, Datta and Lahiri 2000); or use a Bayesian approach (Berger 1985, p. 192, Bell 1999). JLWW propose the jackknife to address both of the two problems.

Before examining how the jackknife addresses the two problems noted, it is worth reminding ourselves of a third problem, which is that use of the MSE result (4) depends on the model being correct. This problem may well be more important and more difficult to address than either of the other two problems. It compromises all the approaches noted to an unknown degree for any particular example.

JLWW's jackknife estimate of MSE is

$$MSE_J = \left\{ g_1(\widehat{A}) - \frac{m-1}{m} \sum_{u=1}^{m} [g_1(\widehat{A}_{-u}) - g_1(\widehat{A})] \right\} + \frac{m-1}{m} \sum_{u=1}^{m} (\widehat{\theta}_{-u} - \widehat{\theta}) \qquad (5)$$

The term in braces estimates $g_1(A)$, with the second term within the braces providing the jackknife bias correction to the plug-in estimate $g_1(\widehat{A})$. The last term in (5) estimates $g_2(A) + g_3(A)$ together, not just $g_3(A)$. These features provide for some generality of the jackknife (e.g., to nonnormal models), though it means that in the context of the linear model ((1),(2)) considered here (5) does not make use of either (i) the asymptotic relation between bias$(g_1(\widehat{A}))$ and $g_3(A)$, or (ii) the exact result for $g_2(A)$. The question arises as to when does the jackknife work better, worse, or about the same as alternatives?

In regard to the question of "How well does the jackknife work?," Jiang *et al.* (2001) report simulation results for some linear and nonlinear (GLIM) models. The jackknife works well in the simulations reported, however, so do all the other approaches considered. In fact, the worst case reported in the simulations of bias in estimated MSE for any method is -10.1% for the naive approach (for a mixed logistic model). This is a relatively small understatement of MSE since, if resulting MSE

estimates were used to construct prediction intervals, the corresponding "standard error" would be understated by only about 5%. Given uncertainties about normality assumptions needed to construct prediction intervals, this amount of understatement of prediction standard error seems relatively unimportant.

In the present paper, the jackknife results JLWW present for the NHIS application look quite reasonable. I decided to examine results from the proposed jackknife approach for an application I am familiar with: estimation of poverty rates of school-aged (5-17 year old) children for the states of the U.S. and DC. These estimates are an important product of the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. The Fay-Herriott model used to produce these estimates is developed in Fay and Train (1997). Bell (1999) discusses Bayesian treatment of this model. Further information on the SAIPE program can be found on the SAIPE web site at http://www.census.gov/hhes/www/saipe.html.

For applying the jackknife to the SAIPE example I used the method-of-moments (MOM) estimator of A used by JLWW in their NHIS example. Table 1 shows these estimates of A for nine years of data to which the model was applied. (1994 is omitted because sampling variances have not been estimated for this year due to complications caused in that year by transition to a redesign of the Current Population Survey which supplies the direct estimates y_i.) Both the not truncated and truncated (at 0) MOM estimates are shown. Maximum likelihood (ML), restricted maximum likelihood (REML), and Bayesian estimates (posterior means) are shown for comparison. These all assume normality, and the Bayesian estimates use flat priors for β and A.

Table 1. Alternative Estimates of A(SAIPE example)

year	ML	REML	Bayes	not truncated MOM	truncated MOM
1989	0	0	1.7	−.1	0
1990	0	0	2.2	1.1	1.1
1991	0	0	1.6	−3.1	0
1992	0	0	1.6	−3.2	0
1993	.4	1.7	3.4	5.8	5.8
1995	0	.2	2.0	.5	.5
1996	0	0	1.9	2.0	2.0
1997	0	0	1.5	−1.3	0
1998*	.7	2.0	3.7	5.8	5.8

*Preliminary results

We notice that the MOM estimates of A are rather unstable. The truncation at zero is frequently required, and for 1993 and 1998 the MOM estimate is quite large relative to the other estimates. ML and REML, though more stable, are not very appealing since these estimates are zero in most years. Bell (1999) notes how estimating A at

zero leads to unreasonable MSE estimates for ML and REML. On the other hand, the Bayesian estimates of A appear much more reasonable (and Bell (1999) notes that resulting Bayesian posterior variances are more reasonable than the frequentist MSE estimates.)

Though I shall omit giving detailed results, it turns out that the jackknife MSE estimates look unreasonable both when A is estimated to be zero and when the MOM estimates of A are large. The poor performance is not generally the fault of the jackknife, however, but simply a result of getting unreasonable estimates of A from MOM. In principle the jackknife could be applied with ML or REML estimation of A, at significantly higher computational cost, though the results in Table 1 suggest this would rarely help in this example. Rather than dwell on possibilities for improving jackknife results by using alternative estimates of A, I will compare jackknife MSE results (using $\widehat{A}_{MOM}$) for two years to illustrate a particular problem that arises when A is estimated at or near zero, and that is more pertinent to the performance of the jackknife.

Notice that when $\widehat{A} = 0$ $g_1(\widehat{A}) = 0$, and all of MSE_J comes from the second and third terms in (5). Tables 2 and 3 below examine the components of MSE_J for two years (1991 and 1989) for which $\widehat{A}_{MOM} = 0$. For both years results are shown for a small number of states for illustration. In the table headings $\widetilde{A}_{MOM}$ denotes the original, not truncated MOM estimates of A from Table 1. Bayesian posterior variances are shown for comparison. It is worth noting that for most states in most years, these posterior variances are very close to what one obtains by substituting the posterior mean of A into $g_1(A)$.

Table 2. Jackknife estimation of MSE for 1991 (SAIPE example)
($\widetilde{A}_{MOM} = -3.1$, $\widehat{A}_{MOM} = \max(0, \widetilde{A}_{MOM}) = 0$)

state	$g_1(\widehat{A})$	$\widetilde{A}_{-u}$	$\widehat{bias}[g_1(\widehat{A})]$	$\widehat{g_2 + g_3}$	**MSE**$_J$	**Bayes**
AL	0	−2.7	0	.7	.7	2.0
AK	0	−2.8	0	.5	.5	2.3
AZ	0	−3.1	0	.5	.5	2.0
AR	0	−3.2	0	1.3	1.3	2.5
CA	0	−2.9	0	.6	.6	1.4
CO	0	−2.8	0	.2	.2	1.6

The $\widetilde{A}_{-u}$ columns in Tables 2 and 3 give the leave-one-out not truncated MOM estimates of A. In 1991 (Table 2) all of the $\widetilde{A}_{-u}$ are negative, with the result that all of the truncated leave-one-out MOM estimates of A ($\widehat{A}_{-u}$) are zero. This is not surprising given that the full sample not truncated MOM estimate for 1991 ($\widetilde{A}_{MOM} = -3.1$) is well below zero—dropping any one observation does not have enough effect to turn any of the $\widetilde{A}_{-u}$ positive. As a result $g_1(\widehat{A}_{-u}) = 0$ for all states u, and since $g_1(\widehat{A}) = 0$ as well, the second term in (5) estimating the bias in $g_1(\widehat{A})$ is zero. Hence,

both terms in the braces in (5) are zero for every state, and MSE_J comes entirely from the third term in (5). This term (labelled $\widehat{g_2 + g_3}$ in the tables) is a jackknife estimate reflecting variation in the small area point estimates $\widehat{\theta}$ due to variation in the leave-one-out estimates of β and A. The resulting MSE estimates tend to look too small both in an absolute sense and relative to the Bayesian estimates—note Colorado (CO) in particular. The MSE estimates also exhibit the same sort of pattern problems noted in Bell (1999) for the ML and REML estimates based on (4). For example, MSE_J for California (CA), despite its large CPS sample, is as high or higher than that for many other states with much smaller samples.

Table 3 shows a different problem that arises for the jackknife estimate of MSE. For 1989 the not truncated MOM estimate of A, $\widetilde{A}_{MOM} = -.1$, is very close to zero. Dropping one observation alters the not truncated MOM estimates as shown in the $\widetilde{A}_{-u}$ column, sometimes yielding positive values, and sometimes yielding negative values. When $\widetilde{A}_{-u} < 0$, $\widehat{A}_{-u}$ is truncated to 0, and $g_1(\widehat{A}_{-u}) = 0$. These states make no contribution to the second term in (5). When $\widetilde{A}_{-u} > 0$, however, $g_1(\widehat{A}_{-u})$ is positive, and these states do contribute to the second term in (5). In fact, since $g_1(\widehat{A}) = 0$ here, the term in braces in (5) is simply minus the sum of these positive terms (multiplied by $(m-1)/m = 50/51$), which turns out to be around 2 for each state. This is the jackknife estimate of bias in $g_1(\widehat{A})$. Subtracting off this bias estimate of around 2 overwhelms the third term in (5), $\widehat{g_2 + g_3}$, resulting in negative estimates of MSE for every state. (The estimates of bias in $g_1(\widehat{A})$ vary only slightly over states since, reintroducing the state subscript i, for state i this term is actually $-\frac{m-1}{m}\sum_{u=1}^{m}[g_{1i}(\widehat{A}_{-u})]$ where $g_{1i}(\widehat{A}_{-u}) = \widehat{A}_{-u}D_i/(\widehat{A}_{-u} + D_i) = \widehat{A}_{-u}/(1 + \widehat{A}_{-u}/D_i) \approx \widehat{A}_{-u}$ since the D_i are much larger than the $\widehat{A}_{-u}$.)

Table 3. Jackknife estimation of MSE for 1989 (SAIPE example)
($\widetilde{A}_{MOM} = -.1$, $\widehat{A}_{MOM} = \max(0, \widetilde{A}_{MOM}) = 0$)

state	g_1	$\widetilde{A}_{-u}$	$\widehat{bias}[g_1]$	$\widehat{g_2 + g_3}$	$\mathbf{MSE}_J$	**Bayes**
AL	0	.16	1.97	.52	−1.4	2.1
AK	0	.04	1.96	.70	−1.2	2.4
AZ	0	−.09	1.97	.37	−1.5	2.1
AR	0	.18	1.97	.71	−1.2	2.4
CA	0	−.03	1.85	.60	−1.2	1.1
CO	0	.10	1.96	.18	−1.7	1.6
CT	0	−1.50	1.96	1.30	−.6	3.0
DE	0	.16	1.97	.38	−1.5	1.8
DC	0	.02	1.98	.90	−1.0	3.3
FL	0	−.13	1.91	.55	−1.3	1.4

The negative estimates of MSE result not just from the poor estimation of A by MOM (though this is a necessary part of the problem), but also from poor estimation

of the bias in $g_1(\widehat{A})$ by the jackknife. This would appear to be a potential problem for the jackknife any time the not truncated estimate of A is very close to zero.

Since the jackknife MSE estimates JLWW present for their NHIS application look quite reasonable, this raises a question about how the SAIPE application differs from the NHIS application. Table 4 provides some answers. It shows, for both surveys, the maximum and minimum values of the estimated signal-to-noise ratio ($\widehat{A}/D_i$) across states, as well as the ratio of the maximum to the minimum state sampling variances ($\max(D_i)/\min(D_i)$).

Table 4. Comparing the NHIS and SAIPE Examples

	$\max\left(\frac{\widehat{A}}{D_i}\right)$	$\min\left(\frac{\widehat{A}}{D_i}\right)$	$\frac{\max(D_i)}{\min(D_i)}$
NHIS	20 (CA)	.1 (SD)	200
SAIPE	1 to 1.5 (CA)	.07 to .1 (DC)	15 to 20

For the NHIS application the signal-to-noise ratio ranges from a very high value of 20 to a very low value of .1. In contrast, for the SAIPE application (for which the results vary some over the years of data) the smallest signal-to-noise ratio is about the same as that for NHIS, but the largest is only around 1 or 1.5. The corresponding ratio of the maximum to minimum sampling variances is 200 for NHIS, reflecting a very wide range of sampling variance across states, but is only 15 or 20 for SAIPE. These data suggest that in the NHIS application the states with large samples provide enough information for reasonably reliable estimation of A (here by MOM), which leads to reasonable looking estimates of MSE by the jackknife (and presumably by other approaches). Small area estimation is needed for those states with small NHIS samples (low signal-to-noise ratios). On the other hand, the CPS direct estimates used in the SAIPE application have sufficiently high levels of sampling error that estimates of A are more unreliable, and conventional frequentist estimates of A frequently run into trouble (as can be seen from Table 1). Resulting estimates of MSE can be unreasonable, and if the not truncated estimate of A is near zero, this can lead to the problem illustrated for the jackknife estimate of MSE.

To generalize the conclusions a bit, estimation of A in the Fay-Herriott model appears to be of more fundamental importance than the choice of alternative approaches to estimating MSE, in the sense that when the data do not provide enough information for reliable estimation of A by conventional frequentist methods, any resulting estimates of MSE are suspect. The Bayesian approach appears to yield more reasonable results in such cases at least by preventing estimates of A near zero. The appeal of the jackknife may be more for cases where nonnormality is a serious concern or the model is nonlinear (e.g., GLIM models), though its performance is still

likely to depend on whether or not the data provide sufficient information for reliable estimation of variances or other dispersion parameters of the model.

References

[1] Bell, William R. (1999) "Accounting for Uncertainty About Variances in Small Area Estimation," *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki.

[2] Berger, James O., (1985) *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

[3] Fay, R. E. and Herriott, R. A. (1979) "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269-277.

[4] Fay, Robert E. and Train, George F. (1997) "Small Domain Methodology for Estimating Income and Poverty Characteristics for States in 1993," American Statistical Association, Proceedings of the Social Statistics Section, 183-188.

[5] Jiang, J., Lahiri, P., and Wan, S. (2001), "A Unified Jackknife Theory," paper submitted for publication.

Can What Partha Lahiri and Company Have Done Help the National Agricultural Statistics Service?

Phillip S. Kott; USDA/NASS

The National Agricultural Statistics Service (NASS) is using component-of-variance small-domain estimation to help estimate the undercoverage in the US Census of Agriculture. The backbone of the Census of Agriculture is an extensive (but not exhaustive) list of farms in the US. Although farms on this list are responsible for over 95% of most agricultural activities, NASS estimates that 13% of all farms operating in 1997 were not on the Census list. One reason for this is that the federal government uses a very liberal definition of a farm: an operation producing at least $1,000 of agricultural output in a year or capable of producing that output.

Leaving aside the merits of the government's definition of a farm, NASS wants the 2002 Census of Agriculture to do a better job than it has in the past providing state aggregates for the farms *not* on the Census list. NASS's key instrument in that endeavor is an area-frame sample designed primarily to estimate the total corn, wheat, soy beans, cotton, and potato acreage and production in the US. A secondary use of this sample is the estimation of aggregates for farms not on NASS's survey and Census lists. For the 2002 Census effort, this sample will be supplemented to better measure the Census-list undercoverage. Nevertheless, NASS believes that only the total numbers of farms missing form the Census list will be reliably estimated at the state level (and even then, certain states like those in New England will have to be combined). For aggregates like the number of missing farms that have horses or the number of missing farms operated by blacks, small-domain techniques will be needed that draw strength from states outside the particular state of interest.

For the purposes of this discussion, let us focus on one particular estimate: the fraction of farms not on the Census list that have a black operator. In truth, NASS will be estimating 20 or 30 fraction like this one, from the fraction of missing farms with horses to the fraction of missing farms with annual sales in a given range. Conceptually, however, they are all the same. One thing to note is that the fractions are many and disparate. Consequently, unlike the problem in Jiang *et al*. (2001), NASS uses a single covariate the fraction of farms on a state's Census list *with* the attribute in question (e.g., a black operator). To begin, we will ignore even that covariate.

Background

Partha Lahiri and his team of collaborators, Jiming Jiang, Shu-Mei Wan, and Chien-Hua Wu, have written a number of papers based on research funded by federal statistical agencies through the National Science Foundation. The question I will address here is whether that research can be of service to NASS in its attempt to estimate the fraction of black-operated farms among those missing from a state's Census list. I begin with some notation borrowing liberally from Jiang *et al*., the particular paper under review. Throughout this discussion, I will refer to the "Lahiri team." That should not be construed as a denigration of Jiang and the other collaborators' contributions.

Let z_{ij} be 1 if a missing farm j in the area sample of State i has a black operator; 0 otherwise. The direct, randomization-based estimator for π_i, the fraction of black-run operations among those missing from the Census list in State i, is

$z_i = \sum_{j \in S(i)} w_{ij} z_{ij} / \sum_{j \in S(i)} w_{ij}$,

where w_{ij} is the sampling weight attached to farm j, and S(i) is the set of farms in the area sample of State i but not on the State's Census list.

That variance of this model under a simple Bernoulli model is

$$\begin{aligned} Var(z_i) &= \{\sum_{j \in S(i)} w_{ij}^2 / [\sum_{j \in S(i)} w_{ij}]^2\} \pi_i (1 \quad \pi_i) \\ &= \pi_i (1 \quad \pi_i)/n_i^*, \end{aligned}$$

where $n^* = \{ [\sum_{j \in S(i)} w_{ij}]^2 / \sum_{j \in S(i)} w_{ij}^2 \}$ is the *effective sample size* in State i. NASS believes that the simple Bernoulli model is appropriate in this context. It ignores the effects of stratification and cluster sampling on variance and assumes that the only role sample weighting plays is in increasing variance and decreasing effective sample size. Nevertheless, weights are used in determining z_i as protection against model failure.

Jiang *et al*. assumes that the D_i can be determined reliably with randomization-based methods. That is not my experience. If sample sizes are not large enough to estimate π_i directly, then estimates of the variance of the direct estimator are even more suspect. Still, nothing is lost if Jiang's randomization-based D_i is replaced by my model-based one.

Suppose we have M "states" for which we need estimates (recall that some states are collapsed together for this purpose). To draw strength from the other states, NASS assumes that each z_i can itself be modeled:

$$\begin{aligned} z_i &= \pi_i + e_i \\ &\pi + v_i + e_i, \end{aligned} \qquad (1)$$

where $E(v_i) = E(e_i) = 0$, $Var(v_i) = A$, and $Var(e_i) = D_i = \pi_i (1 \quad \pi_i)/n_i^*$.

Consider the estimator

$$\begin{aligned} z_i^{(\gamma)} &= (1 \quad \gamma_i) z_i + \gamma_i \sum^M w_k z_k / \sum^M w_k , \\ &= (1 \quad \gamma_i) z_i + \gamma_i z \end{aligned} \qquad (2)$$

where $\gamma_i \approx D_i /(A + D_i)$, w_k is the sum of the sampling weights within S(k), and z is the randomization-based estimator for the fraction of black-run operations among the farms missing from the Census list *nationally*. Since D_i approaches 0 as the effective sample size in i becomes arbitrarily large, $z_i^{(\gamma)}$ is randomization consistent whenever z_i is.

The estimator, $z_i^{(\gamma)}$, is not quite optimal under the component-of-variance model in equation (1). Many (Lahiri and his team included) would not weight the z_k by w_k. Moreover, $z_i^{(\gamma)}$ ignores the variance of z and the covariance between z_i and z. Nevertheless, we will assume for simplicity that M is large enough that such issues hardly matter. More important is the requirement that we estimate A and the D_i before $z_i^{(\gamma)}$ can be operationalized.

To estimate D_i, we need to estimate π_i first, but that is precisely the goal of the entire exercise. It is common to estimate D_i using z_i in place of π_i. Indeed, that is what NASS has been doing for the most part. Note, however, that when $z_i = 0$, D_i must also be zero. This is suspect. Just because we find no black-run operations that not on the Census list in a state area sample does not mean there are no back-run operations missing from the state's Census list anywhere. To avoid

this silliness, NASS has been setting an arbitrary lower bound on its D_i estimate. Nevertheless, the agency's calculation of γ_i remains dependent on a very rickety estimator for π_i.

The variance component A can be estimated using the methods of moments:

$$a^* = [\sum^M z_i^{\,2} \quad (\sum^M z_i)^2/M]/(M \quad 1) \quad [\sum^M z_i(1 \quad z_i)/n_i^*]/M .$$

This formula can produce negative estimates. It is therefore popular to estimate A with

$$a = \max\{0, a^*\}.$$

NASS uses something a bit different:

$$a_{NASS} = \max\{a^*, (1/2)[\sum^M z_i^{\,2} \quad (\sum^M z_i)^2/M]/(M \quad 1)\}.)$$

Although NASS does not think it has the sample sizes to estimate the π_i directly. It does think that it can estimate directly the fraction of black-operated farms among the farms missing from the Census list *nationally* with z. Consequently, if z_i^* denotes NASS's final estimator for π_i, it desires the z_i^* satisfy

$$\sum^M w_i z_i^* / \sum^M w_i \;=\; z .$$

This is the *bookkeeping constraint*.

The Arcsine-root Transform

Jiang *et al*. hits upon a clever way to remove the dependence of the D_i on π_i. Instead of applying the components-of-variance model in equation (1) to the z_i, he applies it to a transform of the z_i:

$$y_i = 2\sin^{-1}(\sqrt{z_i}).$$

(I added a factor of 2 to the transform. It does not effectively change anything, but it makes the arithmetic a bit cleaner.)

One can show that $\mathrm{Var}(y_i) \approx 1/n_i^*$, which is invariant to π_i! Thus, if we replace the z_i in equation (1) by the y_i, the D_i become (nearly) $1/n_i^*$, and the need for early estimates of the π_i is avoided. If NASS were to follow this suggestion, then it would not have to set an arbitrary lower bound on the D_i.

The problem with this transformation is that in invoking it one needs to assume the D_i and A are small. Otherwise, one could not go forward and backward between the original and arcsine-root spaces and preserve near unbiasedness (in particular, the back-transformed solution may not even be unconditionally unbiased for π). If A *is* small, however, then

$$\begin{aligned} D_i &= \pi_i(1 \quad \pi_i)/n^* = (\pi + v_i)(1 \quad \pi \quad v_i)/n_i^* \\ &\approx \pi(1 \quad \pi)/n^* \end{aligned}$$

where the near equality gets better when we take expectations.

This suggests that instead substituting z_i for π_i in $D_i = \pi_i(1 \quad \pi_i)/n_i^*$, NASS begin with $d_i^{(1)} = z(1 \quad z)/n_i^*$, because z is much more stable than z_i, never zero in practice, and fairly close to π_i. The agency can calculate a set of γ_i based on the $d_i^{(1)}$ (and an estimator for A), and then use the computed $z_i^{(\gamma)}$ from equation (2) within $d_i^{(2)} = z_i^{(\gamma)}(1 \quad z_i^{(\gamma)})/n_i^*$. This leads to an iterative process that will likely converge fairly quickly. It should be noted, however, that the Lahiri team's arcsine-root transformation removes the need for iteration.

NASS's Single Covariate and the Bookkeeping Constraint

Unlike in the Lahiri team's formulation, NASS uses a single covariate and no intercept. Instead of the model in equation (1), NASS bases its small-domains estimation on

$$z_i = c_i(\mu + v_i) + e_i,$$

where c_i is the fraction of farms on the Census list in State i that have black operators. How to incorporate this type of information in arcsine-root space is not a trivial question.

Although more intuitively appealing than a model at least half in arcsine-root space (it is unclear whether the c_i should also be so transformed), the model NASS uses has its own conceptual drawback. It is not symmetric in that the model for $1 \quad z_i$ is not linear in $1 \quad c_i$. For fractions like black-operated farms, it is clear that NASS wants to look at z_i rather than $1 \quad z_i$ because it is much smaller. For other fractions, like the fraction of farms with hog production, that is not so straightforward. Indeed, whether z_i or $1 \quad z_i$ is smaller depends upon the state.

Often a simple ratio adjustment is used to enforce the bookkeeping constraint. That is to say, each near-optimal $z_i^{(\gamma)}$ is multiplied by the common factor necessary for the constraint to hold. NASS, however, has been incorporating the constraint directly into the optimality requirement. Rather than minimizing the mean squared error of each state separately. NASS attempts to minimize the weighted sum of the state mean squared errors under the bookkeeping constraint. This would be nearly impossible to do in arcsine-root space.

Variance Estimation

Perhaps the Lahiri team's single largest contribution is in the area of variance estimation, where they propose two simple jackknives to adjust for the asymptotic biases of the conventional variance estimator for the optimal $z_i^{(\gamma)}$, $v_i = D_i a/(a + D_i)$, where a is a method-of-moments estimator for A, and D_i is known.

Even if we accept my model-based formulation, D_i is not known, except approximately in arcsine-root space. More to the point, because of the restriction on a (or, more precisely, a_{NASS}) and the bookkeeping constraint, NASS will not be using a near-optimal $z_i^{(\gamma)}$, although it's estimator can still be put in the form of equation (2) (which I will continue to call $z_i^{(\gamma)}$, without, I hope, undo confusion). With this in mind, it is not clear to me that treating the γ_i NASS uses as fixed in the variance formula:

$$v' = (1 \quad \gamma_i)^2 z_i^{(\gamma)}(1 \quad z_i^{(\gamma)})/n_i^* + \gamma_i^2 a$$

is inappropriate when M is large. When M is less than large, an analogous formula can be derived.

This does not render the Lahiri team's variance formula useless, however. It can still be calculated to estimate just what NASS loses by not using the optimal $z_i^{(\gamma)}$.

What About Confidence Intervals?

Wald confidence intervals for proportions can extend below 0 and beyond 1. Jiang *et al*. point out that the arcsine-root transform appears to be a cure for that.

An alternative is to extend Wilson's (1927) method; e.g., compute a 95% confidence interval by solving the following for π_i :

$$\frac{|z_i^{(\gamma)} - \pi_i|}{\{v'\pi_i(1 - \pi_i)/[z_i^{(\gamma)}(1 - z_i^{(\gamma)})]\}^{1/2}} < 1.96 ,$$

Squaring both sides leads to a easily solvable second degree polynomial in π_i, which can be converted into a asymmetric confidence interval around z_i*.

Concluding Remarks

Although NASS will likely not use the jackknives proposed by Lahiri and his collaborators, I find them extremely useful in principle and remarkably intuitive (where what they extend was not). NASS will also likely not use the arcsine-root transformation. Nevertheless, I think it is fair to say that the exercise of studying what the Lahiri team had done will sharpen what NASS finally does.

Reference

Jiang, J.P., Lahiri, P., Wan S.-M, and Wu, C.-H., Jackknifing in the Fay-Herriot Model with an Example. [earlier in the Proceedings]

Wilson, E.B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, **22**, 209, 212.