Evaluation and Selection of Models for Attrition Nonresponse Adjustment

Eric Slud and Leroy Bailey

Statist. Research Division, US Census Bureau

Methodology Symposium 2009



Eric Slud and Leroy Bailey Evaluation of Nonresponse Adjustments

イロト イポト イヨト イヨト

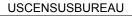
æ

Outline

Longitudinal Nonresponse Weight Adjustment Previous Evaluation Methods Model-based Adjustments **Adjustment Metrics** 2 Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells Computed Results on SIPP 96 3 Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

Definitions motivated by SIPP Survey of Income & Program Participation

- Individuals *i* observed in successive 'Waves'
- Sample s here denotes Wave-1 responders
- Weights w_i are inverse inclusion prob's re-weighted for Wave-1 response & raked.
- Survey items y_i e.g. socSec income indicator are only Wave 1 data values.



・ 回 ト ・ ヨ ト ・ ヨ ト

Definitions motivated by SIPP Survey of Income & Program Participation

- Individuals *i* observed in successive 'Waves'
- Sample s here denotes Wave-1 responders
- Weights w_i are inverse inclusion prob's re-weighted for Wave-1 response & raked.
- Survey items y_i e.g. SocSec income indicator are only Wave 1 data values.
- r_i = indicator of later-wave (say Wave-4) response treated as random beyond probability sample

• \hat{p}_i model-based estimator of $p_i = \Pr(r_i = 1 \mid i \in \mathbf{s})$

USCENSUSBUREAU

くロト (過) (目) (日)

Previous Evaluation Methods Model-based Adjustments

Tracking of time-changes in weights



・ロト ・ 理 ト ・ ヨ ト ・

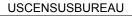
∃ <2 <</p>

Previous Evaluation Methods Model-based Adjustments

Tracking of time-changes in weights

• Dufour et al. 2001 study adjusted weights $w_i r_i / \hat{p}_i$

in simulation using Canadian Labor Force survey.



ヘロト ヘアト ヘビト ヘビト

Previous Evaluation Methods Model-based Adjustments

Tracking of time-changes in weights

 Dufour et al. 2001 study adjusted weights w_ir_i/p̂_i in simulation using Canadian Labor Force survey.

Present Methodology – Internal Evaluation

Target $t_y = \sum_{i \in U} y_i$ wave 1 responder total Estimates initial $\sum_{i \in \mathbf{s}} w_i y_i$ (Horvitz-Thompson)

at later wave, adjusted: $\sum_{i \in \mathbf{S}} w_i y_i r_i / \hat{p}_i$ Adj. Bias $\sum_{i \in \mathbf{S}} w_i y_i \left\{ r_i / \hat{p}_i - 1 \right\}$

SIPP Implemented in Bailey (2004), with SE calculation in Slud & Bailey (2006).

USCENSUSBUREAU

ヘロン 人間 とくほ とくほ とう

Previous Evaluation Methods Model-based Adjustments

Model-based Adjustments

• Cell-based Method: for *i* in cell $C \subset U$,

$$\hat{p}_i = \sum_{j \in C} r_j w_j / \sum_{j \in C} w_j$$

USCENSUSBUREAU

Eric Slud and Leroy Bailey Evaluation of Nonresponse Adjustments

ヘロン ヘアン ヘビン ヘビン

Previous Evaluation Methods Model-based Adjustments

Model-based Adjustments

• Cell-based Method: for *i* in cell $C \subset U$,

$$\hat{p}_i = \sum_{j \in C} r_j w_j / \sum_{j \in C} w_j$$

• Logistic Regression Method: $\hat{p}_i = (1 + e^{-\mathbf{x}_i^{\prime\beta}})^{-1}$ fitted from weighted ML equation, Wave 1 covariates \mathbf{x}_i and later-Wave indicator r_i as response.

USCENSUSBUREAU

ヘロト ヘアト ヘビト ヘビト

Previous Evaluation Methods Model-based Adjustments

Model-based Adjustments

• Cell-based Method: for *i* in cell $C \subset U$,

$$\hat{p}_i = \sum_{j \in C} r_j w_j / \sum_{j \in C} w_j$$

- Logistic Regression Method: $\hat{p}_i = (1 + e^{-\mathbf{x}'_i\beta})^{-1}$ fitted from weighted ML equation, Wave 1 covariates \mathbf{x}_i and later-Wave indicator r_i as response.
- Score-statistic model comparisons: Kim and Kim (2007).

ヘロン ヘアン ヘビン ヘビン

Previous Evaluation Methods Model-based Adjustments

Model-based Adjustments

• Cell-based Method: for *i* in cell $C \subset U$,

$$\hat{p}_i = \sum_{j \in C} r_j w_j / \sum_{j \in C} w_j$$

- Logistic Regression Method: $\hat{p}_i = (1 + e^{-\mathbf{x}'_i \hat{\beta}})^{-1}$ fitted from weighted ML equation, Wave 1 covariates \mathbf{x}_i and later-Wave indicator r_i as response.
- Score-statistic model comparisons: Kim and Kim (2007).
- Model comparisons via Subdomain Bias Estimates

USCENSUSBUREAU

・ロ と ・ 一日 と ・ 日 と ・ 日 と

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

Subdomain Bias & 1st Metric

• Bias over Subdomain \mathcal{D} for k'th survey item

$$\hat{B}_k(\mathcal{D}) = \sum_{i\in\mathcal{D}\cap\mathbf{s}} \left(\frac{r_i}{\hat{p}_i} - 1\right) w_i y_i^{(k)}$$



Eric Slud and Leroy Bailey Evaluation of Nonresponse Adjustments

・ロト ・ 理 ト ・ ヨ ト ・

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

Subdomain Bias & 1st Metric

• Bias over Subdomain \mathcal{D} for k'th survey item

$$\hat{B}_k(\mathcal{D}) = \sum_{i \in \mathcal{D} \cap \mathbf{s}} \left(\frac{r_i}{\hat{p}_i} - 1 \right) w_i y_i^{(k)}$$

• Metric def'n: max Relative Bias over consecutive subsets after random re-ordering $\tau = (\tau(1), \tau(2), \dots, \tau(n))$ of **s**: then averaged over permutations τ :

$$m_k = E_{\tau} \left(\max_{1 \leq a \leq n} |\hat{B}_k(\{\tau(1), \ldots, \tau(a)\})| \right) / \hat{t}_{y^{(k)}}$$

USCENSUSBUREAU

ヘロト ヘアト ヘビト ヘビト

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

Subdomain Bias & 1st Metric

• Bias over Subdomain \mathcal{D} for k'th survey item

$$\hat{B}_k(\mathcal{D}) = \sum_{i \in \mathcal{D} \cap \mathbf{s}} \left(\frac{r_i}{\hat{p}_i} - 1 \right) w_i y_i^{(k)}$$

• Metric def'n: max Relative Bias over consecutive subsets after random re-ordering $\tau = (\tau(1), \tau(2), \dots, \tau(n))$ of **s**: then averaged over permutations τ :

$$m_{k} = E_{\tau} \Big(\max_{1 \le a \le n} |\hat{B}_{k}(\{\tau(1), \ldots, \tau(a)\})| \Big) / \hat{t}_{y^{(k)}}$$

• Monte Carlo Estimate (over permutations τ_c):

$$\hat{m}_{k} = \sum_{c=1}^{R} \max_{1 \le a \le n} |\hat{B}_{k}(\{\tau_{c}(1), \ldots, \tau_{c}(a)\})| / (R \hat{t}_{y^{(k)}})$$

USCENSUSBUREAU

ヘロト 人間 ト ヘヨト ヘヨト

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

Whole-Population Bias vs. Bound

 Random permutations ensure one cannot adjust the average maximum bias to 0 !!

$$|m_{k} - \frac{|\hat{B}_{k}(\mathcal{U})|}{\hat{t}_{y^{(k)}}}| \leq \frac{1.229}{\hat{t}_{y^{(k)}}} \left(\sum_{i \in \mathcal{S}} (\frac{r_{i}}{\hat{p}_{i}} - 1)^{2} (\frac{y_{i}^{(k)}}{\pi_{i}})^{2}\right)^{1/2} = \hat{b}_{k}$$

- Dominant term in m_k or \hat{m}_k , when large $= |\hat{B}_k(\mathcal{U})| / \hat{t}_{\mathcal{Y}^{(k)}}$
- Bound \hat{b}_k on RHS accounts for maximum over subdomains if there is no overall bias.

USCENSUSBUREAU

ヘロン ヘアン ヘビン ヘビン

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

2nd Metric: Preserving Cells

Subdomain biases are most interesting in cells used in current cell-based adjustment:

• Random permutations σ of indices now permute cells A_j , j = 1, ..., J, and individuals within consecutively indexed cells:

$$m_k^* \equiv E_{\sigma} \Big(\max_{1 \leq q \leq n} |\hat{B}_k(\{\sigma(1), \ldots, \sigma(q)\})| \Big) / \hat{t}_{y^{(k)}}$$

USCENSUSBUREAU

ヘロト 人間 ト ヘヨト ヘヨト

Subdomain Bias & 1st Metric Whole-Pop Bias vs. Bound 2nd Metric: Preserving Cells

2nd Metric: Preserving Cells

Subdomain biases are most interesting in cells used in current cell-based adjustment:

Random permutations *σ* of indices now permute cells
 A_j, *j* = 1,..., *J*, and individuals within consecutively indexed cells:

$$m_k^* \equiv E_{\sigma} \Big(\max_{1 \leq q \leq n} |\hat{B}_k(\{\sigma(1), \ldots, \sigma(q)\})| \Big) / \hat{t}_{y^{(k)}}$$

Monte Carlo estimated version of this metric

$$\hat{m}_{k}^{*} \equiv \frac{1}{R} \sum_{r=1}^{R} \max_{1 \le q \le n} |\hat{B}_{k}(\{\sigma_{r}(1), \ldots, \sigma_{r}(q)\})| / \hat{t}_{y^{(k)}}$$

USCENSUSBUREAU

くロト (過) (目) (日)

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

・ロト ・ 同ト ・ ヨト ・ ヨト … ヨ

Adjustment Models Used in SIPP

- 149 Adjustment Cells used in SIPP production: defined in terms of : Education, Income-Level, Labor-force status, Self-employment, Race/Hispanic, Assets, ...
- Logistic regression using variables: Renter, race, `ref-person', education, poverty, some pairwise interactions, and Survey Items (AFDC, SocSec, Unemp, ...)

Wave 4, 12 response fitted separately to Wave 1 predictors.

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

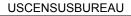
ヘロト 人間 ト ヘヨト ヘヨト

æ

Main Issue. Previously (2006): including Poverty mostly removed whole-population adjustment bias for Poverty.

- Subdomain analyses used to study remaining biases
- add more survey items to logistic model ?

Method based on Metrics.



・ロット (四) ・ (日) ・ (日)

Main Issue. Previously (2006): including Poverty mostly removed whole-population adjustment bias for Poverty.

- Subdomain analyses used to study remaining biases
- add more survey items to logistic model ?

Method based on Metrics.

Items with bad adjustment bias flagged by

$$\hat{m}_k \gg \hat{b}_k$$

• Metric value always of order $\geq \hat{b}_k$.

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

・ロ・ ・ 同・ ・ ヨ・ ・ ヨ・

3

Logistic Models Compared on SIPP 96

	Df	Variables	Dev
Α	7	Wnotsp Renter College RefPer	76558
		Blk Renter*College Blk*College	
В	8	same as A , plus Pov	76545
С	13	same as B , plus Foodst Mdcd	76299
		Heins UnEmp Div	
D	13	same as B , minus Blk*College	76242
		+ Mdcd Heins UnEmp Pov*Heins	
		Mdcd*Heins Heins*College	
Ε	17	same as D + hisp + Famtyp	76017
F	18	C plus Afdc SocSec Emp Mar	76280

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

Model B vs Adjustment Cell

- Logistic Pov biases small, SocSec, Heins, UnEmp large.
- Bounds \hat{b}_k approx. same for all models.

ltem	\hat{m}^{4C}	\hat{m}^{4L}	$\hat{b}_{4,k}$
AFDC	.0067	.0248	.0078
SocSec	.0191	.0116	.0041
Heins	.0085	.0065	.0019
Pov	.0187	.0033	.0047
Emp	.0016	.0017	.0020
UnEmp	.0534	.0594	.0131

USCENSUSBUREAU

・ロト ・ 理 ト ・ ヨ ト ・

ъ

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

・ロト ・ 同ト ・ ヨト ・ ヨト … ヨ

Metric for Models A, D, F, Wave 4

 Successive improvement with more Survey-item predictors.

Item	$\hat{m}^{4,A}$	<i>m</i> ^{4,D}	<i>m</i> ^{4,F}	$b_{4,k}^D$
AFDC	.0175	.0067	.0053	.0077
SocSec	.0117	.0125	.0027	.0041
Pov	.0123	.0032	.0032	.0047
UnEmp	.0626	.0095	.0098	.0139

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

Metrics *m̂*, all Models

- 1st metric, averaged over 12 survey items
- Steady improvement in models with more terms

Model	Wave-4	Wave-12	
Adj.Cell	0.01228	0.04741	
LReg, A	0.01451	0.03942	
LReg, B	0.01504	0.03893	
LReg, C	0.00426	0.02475	
LReg, D	0.00571	0.02812	
LReg, E	0.00481	0.02654	
LReg, F	0.00342	0.00782	

USCENSUSBUREAU

・ロット (四) ・ (日) ・ (日)

ъ

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

イロト イポト イヨト イヨト

Metrics \hat{m}_k^* Across Models

- 2nd metric, by single survey item
- Richer models often slightly better,
 but now Models C and F virtually tied !!

item	ModB	ModC	ModD	ModF	Adj.Cell
					.0697
SocS	.0300	.0305	.0310	.0287	.0332
Pov	.0632	.0574	.0576	.0566	.0572
Emp	.0281	.0263	.0261	.0259	.0227

Adjustment Models Model Comparisons Comparison using Raked vs Raw Weights

イロト イポト イヨト イヨト

Metrics \hat{m}_k^* Across Models

- 2nd metric, by single survey item
- Richer models often slightly better,
 but now Models C and F virtually tied !!

item	ModB	ModC	ModD	ModF	Adj.Cell
					.0697
SocS	.0300	.0305	.0310	.0287	.0332
Pov	.0632	.0574	.0576	.0566	.0572
Emp	.0281	.0263	.0261	.0259	.0227

• Low-dimensional Model C as good as Adjustment Cells.

くロト (過) (目) (日)

Metrics in Wave 4 with/without Raking Weights

- Raking was done in SIPP 1996 to 126 demographic cells.
- Compare metric values and models, in Wave 4, with and without raking.

Metric	Raked	ModD	ModF	Modl	ModIII	AdjCel
m _k	No	.0057	.0034		.0039	.0123
	Yes	.0047	.0057	.0046	.0045	.0083
m_k^*	No	.0065	.0044	.0046	.0045	.0127
	Yes	.0046	.0056	.0046	.0045	.0082

ヘロト ヘワト ヘビト ヘビト

Metrics in Wave 4 with/without Raking Weights

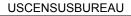
- Raking was done in SIPP 1996 to 126 demographic cells.
- Compare metric values and models, in Wave 4, with and without raking.

Metric	Raked	ModD	ModF	Modl	ModIII	AdjCel
m _k	No	.0057	.0034	.0039	.0039	.0123
	Yes	.0047	.0057	.0046	.0045	.0083
m_k^*	No	.0065	.0044	.0046	.0045	.0127
	Yes	.0046	.0056	.0046	.0045	.0082

• Raking makes little difference with best models (F,I,III), but helps a lot with poor ones (D, AdjCel).

Summary

- Internal evaluation of attrition nonresponse adjustment.
- Introduced 2 metrics for largest absolute relative bias over subsequence in random ordering of survey items.



★ E → ★ E →

< 🗇 🕨

Summary

- Internal evaluation of attrition nonresponse adjustment.
- Introduced 2 metrics for largest absolute relative bias over subsequence in random ordering of survey items.
- Metrics prevent 'adjusting away' biases at whole-pop level.
- Metrics rate models differently from Deviance: 1st favors greater adjustment. 2nd Metric favors less.

・ 同 ト ・ ヨ ト ・ ヨ ト

Summary

- Internal evaluation of attrition nonresponse adjustment.
- Introduced 2 metrics for largest absolute relative bias over subsequence in random ordering of survey items.
- Metrics prevent 'adjusting away' biases at whole-pop level.
- Metrics rate models differently from Deviance: 1st favors greater adjustment. 2nd Metric favors less.
- Re-did metric model-comparisons in SIPP after implementing population-control '2nd stage adjustment' (Raking).

USCENSUSBUREAU

References

Dufour, J., Gagnon, F., Morin, Y., Renaud, M. and Särndal, C.-E. (2001) Survey Methodology 27, 97-108.

Bailey, Leroy (2004)

Census Bureau internal report, Nov. 16, 2004.

- Slud, E. and Bailey, L. (2006) Estimation of attrition biases in SIPP. ASA Surv. Res. Meth. Proceedings, JSM 2006.
- Kim, Jae-Kwang and Kim, Jay (2007) Canadian Journal of Statistics, to appear.

USCENSUSBUREAU

< 回 > < 回 > < 回