Resampling Methods for Sample Surveys

A. J. Canty

July 2, 1998

Acknowledgements:
Anthony Davison, EPFL; Dave Elliot, ONS;
Rudi Peters and Beat Hulliger, SFSO.
Data supplied by the Data Archive.

Labour Force Surveys

- Large sample size.
- Panel Surveys.
- Usually 20% renewal each period.
- Stratified cluster sampling.
- Sample-based weights assigned to each individual.
- Iterative Proportional Fitting (IPF) used.
- Estimates of change are important.

Aims of the Project

1. Compare resampling-based approaches to standard error calculation with existing methods.
2. Examine how to take the randomness of the weights into account.
3. Calculation of standard errors for change estimates.
4. Implementation issues for the new methods.

Estimates

$$\begin{aligned}
\widehat{\theta}_{tot}(y) &= \sum_{h,j,k} w_{hjk} y_{hjk}, \\
\widehat{\theta}_{av}(y) &= \frac{\sum_{h,j,k} w_{hjk} y_{hjk}}{\sum_{h,j,k} w_{hjk}}, \\
\widehat{\theta}_{rat}(y,z) &= \frac{\sum_{h,j,k} w_{hjk} y_{hjk}}{\sum_{h,j,k} w_{hjk} z_{hjk}}.
\end{aligned}$$

Existing Standard Errors

$$\begin{aligned} \mathrm{se}(\widehat{\theta}_{av}) &\approx d(Y)\left\{\widehat{\theta}_{av}(1-\widehat{\theta}_{av})/n\right\}^{1/2}, \\ \mathrm{se}(\widehat{\theta}_{tot}) &\approx N\mathrm{se}(\widehat{\theta}_{av}) \end{aligned}$$

- $d(Y)$ is the Design Effect for the variable Y.
- Delta method used with the above and a covariance estimate for ratios.
- For estimates of a change $\widehat{\theta}_1 - \widehat{\theta}_2$ we have
  $\mathrm{var}(\widehat{\theta}_1 - \widehat{\theta}_2) \approx (1-rk)\left\{\mathrm{var}(\widehat{\theta}_1) + \mathrm{var}(\widehat{\theta}_2)\right\}$
  where
  - k is the sample overlap,
  - r is the auto-correlation at lag one period.

Resampling-based Methods

1. Jackknife
2. Jackknife Linearization
3. Balanced Repeated Replication
4. Bootstrap

Jackknife

Delete each cluster in turn and find $\widehat{\theta}_{-hj}$.

$$v_J(\widehat{\theta}) = \sum_{h=1}^{H} \frac{(1-f_h)(n_h-1)}{n_h} \sum_{j=1}^{n_h} (\widehat{\theta}_{-hj} - \bar{\widehat{\theta}}_{-h})^2,$$

- Theory says weights must be recalculated for each replicate.
- Delete groups of clusters within strata.
- Delete across all strata simultaneously.

Jackknife Linearization

- Replaces replicates with analytic derivatives.
- More complicated for estimates of change.
- Depends on the particular statistic of interest.

Balanced Repeated Replication

- Half-sampling assumes there are two clusters in each stratum.
- In replicates only one of the two clusters are kept.
- Requires $L \geq H + 1$ balanced replicates.
- For $n_h > 2$, make two groups of clusters per stratum and treat groups as clusters.
- Weights must be adjusted for each half-sample.

$$v_{BRR}(\widehat{\theta}) = L^{-1} \sum_{l=1}^{L} (\widehat{\theta}_l^{\dagger} - \widehat{\theta})^2.$$

Bootstrap

- Use the sample to estimate the population distribution.
- Take samples from this population in the same way as the original.
- For small f_h, can just take observations with replacement.
- Bootstrap samples are reweighted to the original margins.
- For bias and variance estimation $R = 100$ replicates is sufficient.

$$v_B(\hat{\theta}) \;=\; (R-1)^{-1}\sum_{r=1}^{R}(\hat{\theta}_r^* - \bar{\hat{\theta}}^*)^2,$$

Simulation Study

- Based on the UK LFS of autumn and winter 1995.
- 1250 addresses stratified by 10 regions.
- Overlap maintained in samples.
- Strata (Region × Month) about the same size as for LFS.
- Calibration to three margins similar to the those used in the LFS.
- Total number unemployed.

Target Values

- Repeated sampling variance from 10,000 samples.

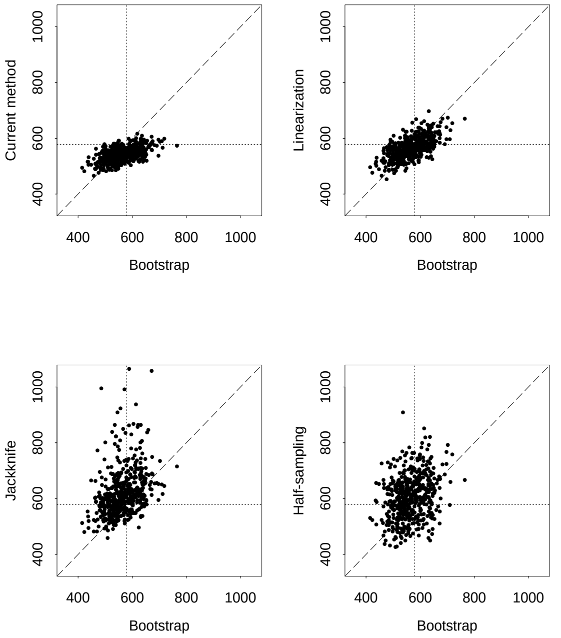
Comparison of Methods

- 500 samples generated.
- Design effect = 1.05, r = 0.6, k = 0.8.
- 3 levels of reweighting used
- For bootstrap and BRR each of the 30 strata is divided into 3 sub-strata to give H = 90.
- L = 96 replicates for BRR.
- R = 100 replicates used for bootstrap.
- Each of the 30 strata divided into 10 groups for jackknife.

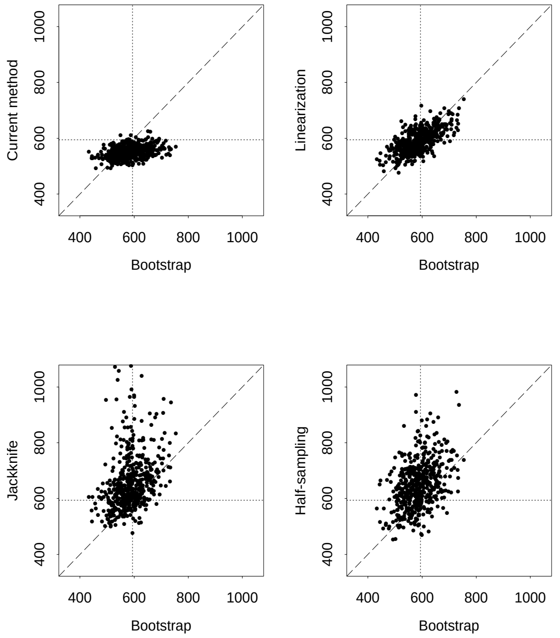
Percentage relative bias (RB), standard deviation (SD) and mean square error (MSE) of standard errors for total unemployed. Full reweighting used where relevant.

Statistic	Method	RB	SD	MSE
Total	current	−6.1	25.4	1880
	bootstrap	−2.5	55.0	3230
	jackknife	7.8	96.8	11400
	linearization	−2.4	38.4	1650
	BRR	4.8	79.6	7100
Change	current	−7.2	21.6	2290
	bootstrap	−1.4	55.5	3140
	jackknife	10.8	107.0	15600
	linearization	−0.8	41.8	1770
	BRR	9.3	85.5	10400

Standard errors for 500 estimates of total unemployed



Standard errors for 500 estimates of change in total unemployed



Results of simulation study

- Current method very stable but biased downwards.
- BRR and jackknife quite unstable.
- Bootstrap and jackknife linearization comparable and provide the best estimates overall.
- Results comparable for other statistics.

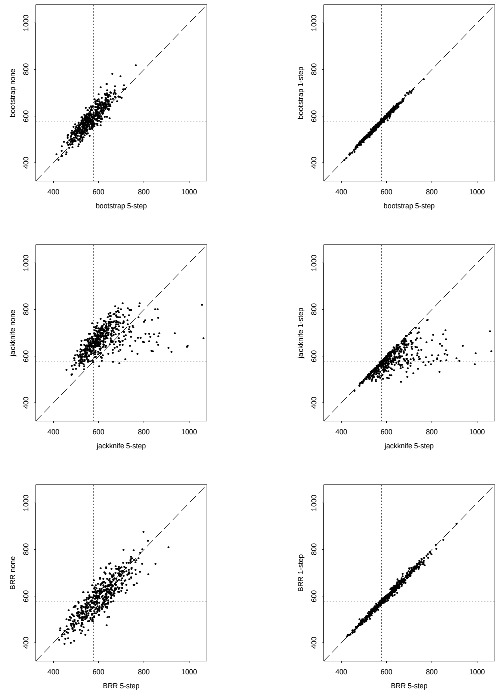
Effect of Reweighting

- One step reweighting is sufficient in all cases.
- For the jackknife one-step reweighting essential for static estimation.
- Five-step reweighting in jackknife can harm performance.
- For BRR and bootstrap, there is little effect.

Percentage relative bias (RB), standard deviation (SD) and mean square error (MSE) of standard errors for total unemployed with different levels of reweighting.

Method	Reweighting	RB	SD	MSE
bootstrap	none	1.3	61.4	3820
	1-step	−2.7	54.6	3220
	5-step	−2.5	55.0	3230
jackknife	none	15.8	58.3	11800
	1-step	0.3	50.1	2500
	5-step	7.8	96.8	11400
BRR	none	2.8	81.5	6880
	1-step	4.1	78.4	6700
	5-step	4.8	79.6	7100

Standard errors for 500 estimates of total unemployed with different levels of reweighting.



Percentage relative bias (RB), standard deviation (SD) and mean square error (MSE) of standard errors for change in total unemployed with different levels of reweighting.

Method	Reweighting	RB	SD	MSE
bootstrap	none	0.3	58.8	3460
	1-step	−1.7	55.0	3120
	5-step	−1.4	55.5	2140
jackknife	none	0.0	51.2	2610
	1-step	1.9	53.6	2990
	5-step	10.8	107.0	15600
BRR	none	3.8	80.8	7030
	1-step	8.2	83.6	9370
	5-step	9.3	85.5	10400

Standard errors for 500 estimates of change in total unemployed with different levels of reweighting.