Computation of the variance of longitudinal estimations for the swiss survey on added value

Lionel Qualité & Yves Tillé

University of Neuchâtel

April 2005

Sampling design

Aim : estimate the added value produced in Switzerland and its evolution.

Fixed sample from 1999 to 2001, stratified by activity and size of of company, as recorded in the last business census.

Big companies are all part of the sample.

Composition: 11200 companies, 6900 of which are big, 1000 medium sized and 3300 small companies. They are dispatched in 46 activity strata.

Observed sample

Average response rate 52%.

Higher reponse rate for big companies: 66% of them answered when 45% of medium sized and only 28% of small ones did.

Due to this non-response the observed sample is not the same during the three years (involuntary rotation of 10 to 15% of the sample).

We do not know if a company that didn't answer has disappeared or if it is a true non-response.

Estimator used by the SFSO

Post-stratified estimation by type of activity as declared in the survey (coincides mostly with the sampling strata)

Creation of a 'surprise poststratum' of huge companies : the 5% largest companies in the observed sample. They are treated separately, receive a sampling weight equal to 1 and it is thereafter assumed that all companies of such size have answered.

Robustification for remaining companies: a multiplicative factor for the weight of the sampled companies is computed. It's equal to 1 for most companies and between 0 and 1 for outliers.

Ratio estimator using number of employees within poststrata.

Variance of transversal estimators

Huge companies are not included in the computation, we suppose they do not bring variance to the estimation

Small strata are collapsed (usually small companies with medium companies in a same activity stratum)

Hypothesis: uniform nonresponse within strata

Calibration taken into account with a regression residue technique on robustified variables (see Deville 1999)

Longitudinal estimation

Basic estimator: $\widehat{Y}_{t+1} - \widehat{Y}_t$ or $\widehat{Y}_{t+1}/\widehat{Y}_t$.

Modelisation: bidimensionnal stratified simple random sampling with fixed sizes $|s_{t+1} \backslash s_t|, |s_{t+1} \cap s_t|, |s_t \backslash s_{t+1}|$

Variance:

$$V(\widehat{Y}_{t+1} - \widehat{Y}_t) = V(\widehat{Y}_{t+1}) + V(\widehat{Y}_t) - 2Cov(\widehat{Y}_{t+1}, \widehat{Y}_t),$$

where

$$Cov(\widehat{Y}_{t+1}, \widehat{Y}_t) = N^2 \left(\frac{|s_t \cap s_{t+1}|}{|s_t||s_{t+1}|} - \frac{1}{N} \right) S_{y_{t+1}, y_t}.$$

Specificities

Use of regression residuals at t and $t+1$ and robustified variables.

Few companies change their activity.

But rotation in the poststratum of huge companies is not negligible (around 15%).

Huge companies

- First option: if a company is in the stratum of huge companies at time t OR $t+1$, it does not participate in covariance (we accept the census hypothesis).
- Second option: only the companies that are 'huge' at t AND $t+1$ are kept in the poststratum of huge companies and the others are returned to their original stratum.

Estimation of the covariance

Two basic possibilities:

- Estimate S_{y_{t+1},y_t} on the intersection of the samples.
- Estimate the correlation on the intersection of the samples and multiply it by $\sqrt{s^2_{y_{t+1}} \cdot s^2_{y_t}}$, where $s^2_{y_t}$ (resp. $s^2_{y_{t+1}}$) is computed on the whole sample s_t (resp. s_{t+1})

Since the variances and covariances are computed on different samples it may happen that we end up with negative estimates in some strata.

For further discussion see berger (2004)

Evolution of the production value

NOGA	VP00-VP99	ET Indep.	ET Uni	Gain (in%)
45	3305854	2354570	872321	63
74	-776044	4383539	1981214	55
51	3069273	2114168	939151	56
24	4333886	1101315	1002355	09
52	-96368	810388	527083	35

Evolution of the added value

NOGA	VA00-VA99	ET Indep.	ET Uni	Gain (in%)
45	1958965	913688	316500	65
74	684613	2995872	1043334	65
51	1903233	1473295	723266	51
24	364955	472427	449355	05
52	-359760	589773	430257	27

Comparison of the different options

Standard error for the added value

NOGA	ET Uni	ET S4 mini	ET Corr
45	316500	316742	433205
74	1043334	1061718	1560833
51	723266	739532	809785
24	449355	557840	423752
52	430257	458820	381861

Large differences in some strata.

Using only the common part of the samples

Standard error for the production value and the added value

	VP			VA		
Noga	ET Uni	ET Inter	Gain (in %)	ET Uni	ET Inter	Gain (in %)
45	872321	616743	29	316500	301468	05
74	1981214	1028663	48	1043334	674998	35
51	939151	625945	33	723266	425322	41
24	1002355	203061	80	449355	73189	84
52	527083	382063	28	430257	311259	28

... however the evolution estimations do not match with the published totals.

Best linear combination

We can look for the unbiased estimator with minimal variance in the family:

$$\left\{\alpha \widehat{Y}_{s_{t+1}\setminus s_t} + \beta \widehat{Y}_{s_{t+1}\cap s_t} + \gamma \widehat{Y}_{s_t\cap s_{t+1}} + \delta \widehat{Y}_{s_t\setminus s_{t+1}}\right\},\ (\alpha,\beta,\gamma,\delta) \in \mathbb{R}^4$$

A sufficient condition to have an ubiased estimator is that $\alpha + \beta = 1$ and $\gamma + \delta = -1$.

References

Berger, Y. (2004).Variance Estimation for Measures of Change in Probability Sampling. Canadian Journal of Statistics, vol 32.

Caron, N. and Ravalet, P. (2000). Estimation dans les enquêtes répétées: Application à l'enquête emploi en continu. Technical Report 0005, Méthodologie Statistique, INSEE, Paris.

Deville, J.-C. (1999). Variance estimation for complex statistics ans estimators: linearization and residual techniques. *Techniques d'enquête, Survey methodology*, 25:193–204.

Hulliger, B. (1999). Simple and robust estimators for sampling. In *Proceedings of the Section on Survey Research Methods*, pages 54–63. American Statistical Association.

Peters, R., Renfer, J.-P., and Hulliger, B. (2001). Statistique de la valeur ajoutée : procédure d'extrapolation des données. Technical report, Office fédéral de la Statistique.