On pooling of data and measures

Vijay Verma, Francesca Gagliardi, Caterina Ferretti

Working Paper n. 84, November 2009

On pooling of data and measures[1]

Vijay Verma (Università di Siena), Francesca Gagliardi (Università di Siena),

Caterina Ferretti (Università di Firenze)

verma@unisi.it, gagliardi10@unisi.it, ferretti@ds.unifi.it

Abstract

Pooling of data means statistical analysis using multiple data sources relating to multiple populations. It encompasses averaging, comparisons and common interpretations of the information. Different scenarios and issues also arise depending on whether the data sources and populations involved are same/similar or different. This paper is primarily concerned with cumulation over space and time from repeated multi-country surveys, taking illustrations from two major European social surveys. Simple model are developed to illustrate the effect on variance of pooling over correlated samples, such as over waves in a rotational panel design.

1. Introduction: aspects of pooling

1.1 Objectives

By pooling we mean statistical analysis or the production of estimates on the basis of multiple data sources, possibly relating to multiple populations. There are three fundamental objectives of pooling of statistical data or estimates.

(1) Cumulation or aggregation in order to obtain more precise estimates, albeit normally with some loss of detail.

(2) Comparisons of trends and differences across populations and times, for instance comparisons between different populations, between different geographical parts or times for a given population.

[1] This work has been conducted under the project SAMPLE – Small Area Methods for Poverty and Living Conditions Estimates - European Union 7th Framework Program - Project/Contract No: EU – FP7 - SSH-2007-1 Grant Agreement no 217565 coordinated by Prof. Monica Pratesi, and under the Italian PRIN research project n. 2007HEWTBE_003 coordinated at National level by Prof. G.M. Giorgi and locally carried out by the C.R.I.DI.RE. research centre.

(3) Meeting the more general and broader objective of *common interpretation* of statistical information from different sources and/or for different populations in relation to each other and against common standards.

1.2 Prerequisite: comparability

For meaningful pooling whether of micro data or of estimates, it is necessary that the different data sources are "comparable". The concept of comparability implies the requirement that data or estimates can be legitimately, i.e. in a statistically valid way, put together (aggregated, pooled), compared (differenced), and interpreted (given meaning) in relation to each other and against some common standard. Comparability is absolutely central to the problems and procedures of pooling of data and estimates. Comparability is a matter of degree. A "sufficient" degree of comparability is a precondition for such pooling to be meaningful (Verma, 2002).

1.3 Diverse scenarios

Procedures and problems in pooling depend on whether the population and the sample involved in the pooling are *similar* (or same) or are *different* for the different element being pooled. How similar or different the sources are is actually a matter of degree: there is no simple dichotomy "same" versus "different". At the one extreme, we have the situation where both the population and the sample (or other types of data sources) involved are different: the data or estimates are being pooled across different population, using different sources of data in each. At the other extreme, we have the situation where both the population and the sample are the same or similar.

On this basis, we may distinguish four main types of situations or scenarios; within each scenario further subtypes may be identified. The important point is that the following distinctions are not necessarily sharp: being the "same" or "different" is a matter of degree.

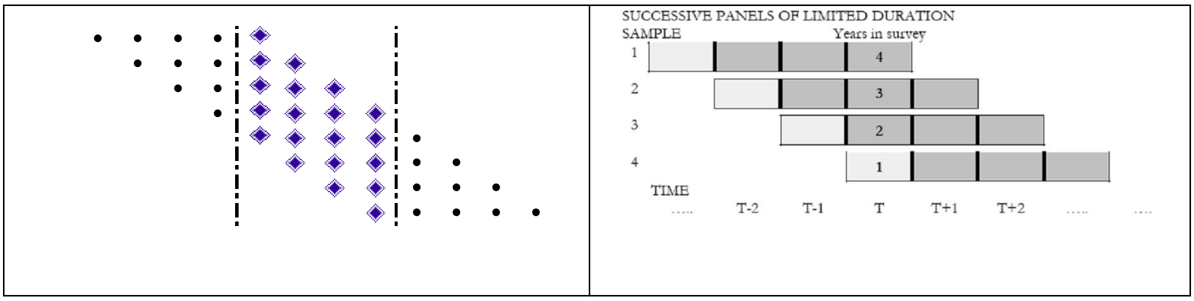
	Data source	
Population	same/similar (s)	different /dissimilar (d)
Same/Similar (S)	S.s	S.d
Different /Dissimilar (D)	D.s	D.d

Issues of comparability are more severe when the data sources are different, and specially when the populations involved are also different. An example of scenario (D.d) is the Luxembourg Income Study (LIS), which uses different sources in different countries for constructing a data source for comparative research on income distribution. Scenario (S.d) means using different sources to obtain a more complete picture for a given population, such as from income and expenditure surveys. More often, we are dealing with pooling of data from similar sources. Typically scenario (D.s) involves pooling over space (e.g. over countries in a multinational survey), and scenario (S.s) involves pooling over time (e.g. in a periodic survey). It is these latter scenarios that we address in this paper.

2 Illustrations from European social surveys

The two most important regular social surveys in the EU are the Labour Force Survey (EU-LFS) and Statistics on Income and Living Conditions (EU-SILC). The EU-LFS was initiated at EU level in 1960, with a systematic common framework adopted from 1983. It is a large sample survey, conducted in all EU countries on a continuous basis, providing quarterly and annual results on labour participation along with socio-demographic and educational variables. Annually ad-hoc modules dedicated to specific topics supplement the core survey. The EU-SILC was launched starting from 2003 in some countries; it covered 27 EU and EFTA countries by 2005, and all 30 by 2008. In each country it involves an annual survey with a rotational panel design. Its content is comprehensive, focusing on income, poverty and living conditions.

Both EU-LFS and EU-SILC involve comprehensiveness in the substantive dimension (coverage of different topics), in space (coverage of different countries), and in time (regular waves or rounds). EU-SILC is stronger in the substantive dimension, and EU-LFS in the spatial (providing reliable estimates at the regional level given its large sample size). Both are strong in the time dimension: EU-LFS providing regular and frequent estimates of levels and net changes, and EU-SILC providing longitudinal indicators (such as persistent poverty) at the micro level. EU-LFS involves diverse types of rotational designs; a simple and common one is illustrated below on the left hand side. In this example, a sample address stays in the survey for 5 consecutive quarters before being dropped. The subsamples contributing to a particular year have been identified in the central part of the diagram. By contrast, for EU-SILC most countries use the standard rotational household panel design shown below on the right. Here the survey is annual, and each panel stays in the survey for four consecutive years.



In the following sections, we discuss a some selected technical issues.

3. Pooling of data versus pooling of estimates

We may also distinguish between *pooling of data*, i.e. aggregation of micro-level data for the same or different populations, surveys and times, on the one hand, and the *pooling of estimates*, i.e. the production of a common estimate as a function of estimates produced from individual data sources.

Let us consider estimate ϕ_i for a certain statistic for country *i*. In comparisons, each ϕ_i of course receives the same weight. For estimates aggregated over EU countries, of the form $\phi = \Sigma P_i.\phi_i$, the most common practice by far is to take the weights P_i in proportion to the countries' population size, thus producing statistics for the 'average EU citizen'.

By contrast, in much policy debate, it is the situation in the 'average EU country' that is of interest; this amounts to taking the P_i values as equal. But it can also be argued that countries as well as individual citizens are both relevant as units, so that larger countries could be given more weight, but less than proportionate to their population size (Verma, 1999).

Whatever the choice of P_i, the above formulation involves pooling country-level estimates. Given standardised data sets from all countries, such as in EU-LFS or EU-SILC, pooling at the micro-level is also possible, with unit weights w_{ij} scaled as $w'_{ij} = w_{ij}.\left(P_i / \Sigma w_{ij}\right)$.

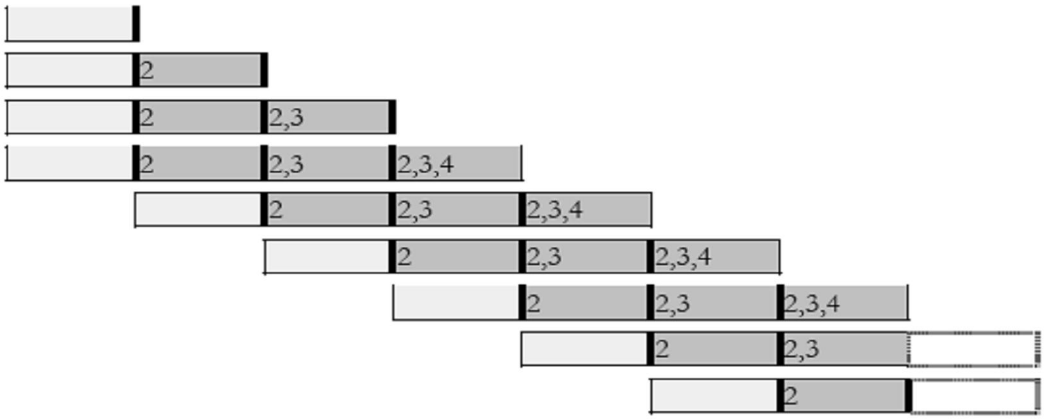
For ratios of the form $\phi_i = \Sigma w_{ij}.v_{ij} / \Sigma w_{ij}.u_{ij}$, the macro and micro pooling give the same result, except that the former is similar to a 'separate' and the latter to a 'combined' type of ratio estimate. In the above form, the contribution of any unit *j* to the estimate does not depend on the values of other units (*k*) in the sample. It does for some other statistics, e.g. measures of income distribution: median income, disparity (Gini), poverty rates, etc. Here the useful distinction is whether the dependence is on units only within a population, or on all units in the pooled populations. Conventional poverty measures using the national poverty line are an example of the former; here pooling across countries is essentially *macro-level*. By contrast, for poverty measures with reference to a common EU poverty line, the pooling across countries has to be at the *micro-level*.

4. Panels in a rotational design

In a rotational design such as of EU-SILC, each cross-section is made up of a number of panels or subsamples (see the figure below). In computing measures for the cross-section, the common practice is simply to pool the cases from the subsamples. This amounts to giving each subsample a 'weight' in proportion to its sample size, i.e. inversely proportion to its expected variance, which is an efficient (optimal) procedure in the absence of bias. However, in a rotational design the subsamples are of different ages, and the older ones can be expected to be more biased, both because of changes in the population since they were first selected, and due to selective non-response.

Consider two panels with same variance V^2, but the second (older) one also subject to bias B. Pooling them with weights W_1, W_2 respectively ($W_1+W_2=1$) gives MSE composed of variance $V^2.(W_1^2 + W_2^2)$ and bias^2 $W_2^2.B^2$. Variance is minimised with $W_1=W_2=0.5$, but bias can be reduced by taking $W_2<0.5$, i.e. giving less weight to the older panel. The optimal choice of the weights depends on the bias ratio B^2/V^2.

CUMULATION OF LONGITUDINAL OBSERVATIONS



CUMULATIVE NUMBER OF LONGITUDINAL OBSERVATIONS

1. SUBSAMPLES PROVIDING CROSS-SECTIONAL DATA								
4	8	12	16	20	24	28	…..	4*Y
2. SUBSAMPLES PROVIDING YEAR-TO-YEAR TRANSITIONS								
0	3	6	9	12	15	18	…..	3*(Y-1)
3. SUBSAMPLES PROVIDING 3 YEARS LONGITUDINAL OBSERVATIONS								
0	0	2	4	6	8	10	…..	2*(Y-2)
4. SUBSAMPLES PROVIDING 4 YEARS LONGITUDINAL OBSERVATIONS								
0	0	0	1	2	3	4	…..	(Y-3)
SURVEY YEAR								
1	2	3	4	5	6	7	…..	Y

Note: The numbers in the cells of the diagram indicate the type(s) of observations provided by the subsample. The above numbers may be multiplied by the subsamples size to obtain the cumulated number of observations.

5. Poverty rates with poverty lines at different thresholds and levels

5.1 Poverty lines at different thresholds

In the standard analysis, as for instance in Laeken indicators, the poverty line is defined as a certain percentage (*threshold* x%) of the median income of the national population. By ‘poverty line *threshold’* we mean the percentage of the median income defining the poverty line. Different values of the poverty rate are obtained depending on the threshold (i.e. on ‘x’) of the chosen poverty line. The Laeken set at the national level includes a measure of dispersion around the at-risk-of-poverty threshold, computing the percentage of persons, over the total population, with an equivalised disposable income below, respectively, 40%, 50%, 60% and 70% of the national median equivalised disposable income, 60% being the main threshold. The substantive objective of introducing indicators of dispersion around the poverty line is to take more fully into account differences among countries in the shape at the lower end of the income distribution. Lower thresholds isolate the more severely poor and tend to be more sensitive in distinguishing among countries or other population groups being compared. As the threshold is raised, this sensitivity generally tends to fall: clearly in the extreme case when 'x' is taken as 100% (poverty line equal to the median), the poverty rate in all situations is 50%, by definition.

In addition to the above systematic differences, the results from using different poverty line thresholds are also likely to be affected by irregularities in the empirical income distribution. Irregularities are larger when the distributions are estimated from smaller samples, as normally is the case for disaggregated estimates by region. It is this consideration which is likely to dominate in the context of constructing regional measures.

In view of the reduced sample sizes in moving to the regional level, it is generally desirable to avoid producing too many individual figures each subject to large sampling variability. Instead, it would seem a good idea to compute poverty rates with reference to several different thresholds, but then to consolidate them, such as by taking an appropriately weighted average of them for comparisons across regions. In specific terms, a single measure based on suitable consolidation over, say, 50%, 60% and 70% of median poverty lines, would be preferable to separate indicators for each of these level such as in the Laeken Indicators list.

There are also substantive considerations in such consolidation. The rate consolidated over different thresholds provides a summary or overall measure of different degrees of severity of poverty contained within the given income distribution (Verma *et al.*, 2005).

5.2 Poverty lines at different levels

The *level* of poverty line indicates the population level at which the income distribution is pooled for the purpose of defining the poverty line. Commonly used poverty-related indicators are based on country poverty lines determined on the basis of the national income distribution. It is necessary to consider other levels of the poverty line: lines defined from income distribution pooled across countries, as well as regional poverty lines at various levels (such NUTS1, NUTS2, …) within the country. For instance, with increasing integration in the European Union, and increasing need for European level analysis, the use of EU-wide benchmarks becomes increasingly justified and rewarding. Similarly, indicators based on regional poverty lines become more important with devolution of social inclusion policies. The use of different poverty lines brings out different aspects. Poverty lines defined at different levels capture different mixes of ‘relative’ measures (concerning purely income *distribution*) and ‘absolute’ measures (concerning differences in income *levels*). Pooling (with some appropriate weights) measures computed with different poverty line levels amount to consolidating the diverse measures into a single (or fewer) measures taking into account multiple dimensions of the situation.

5.3 Averaging over different types of measures

The above argument can also be extended to averaging over different types of statistics, not just poverty rates, each capturing a different dimension or aspect. Such averaging will require appropriate rescaling of the different types of measures before they can be pooled together to provide more ‘consolidated’ measures.

5.4 Aggregation over time (waves)

Where the information comes from sample surveys of limited size, a trade-off is required between temporal detail and geographical breakdown. In order to achieve greater geographical disaggregation (e.g. by region), the emphasis has to be shifted away from the study of trends over time and longitudinal measures to essentially cross-sectional measures aggregated over suitable time periods, so as to illuminate the more stable aspects of the patterns of variation across regions. Average of wave-specific poverty rates over waves provides an indicator reflecting the overall situation over the period covered. The measures constructed from averaging over waves tend to be more robust than results based only on one wave. They increase precision, help to smooth out short-term trends and bring out more clearly the underlying structural relationships of interest.

6. Effect of pooling on variance

We use EU-SILC standard ample structure to illustrate the effect of aggregation on resulting variance. In this design, each cross-section consists of four panels or subsamples, introduced one by one over the preceding years.

6.1 Reduction in variance by pooling data for subsamples

In aggregating over subsamples, variance decreases in inverse proportion to sample size, provided that the subsamples making up the total sample are independent. This is the case with EU-SILC samples where each subsample is based on a different set of clusters. There are also a number of designs in which different subsamples involve different households but all from a common set of clusters. Here the design effects tends to increase as the subsamples are pooled, so that the gain in precision is smaller than proportionate to sample size.

6.2 Reduction of variance from averaging different poverty thresholds

As noted, some gain in sampling precision can be obtained by computing poverty rates using different thresholds, and then taking their weighted average using some appropriate pre-specified (i.e., constant or external) weights. A quantitative indication of the magnitude of this gain may be obtained on the following lines.

Consider three poverty line thresholds, with poverty rates p_i, $p_1 < p_2 < p_3$, and that with fixed weights W_i, $\Sigma W_i = 1$, a consolidated rate is computed as $p = \Sigma W_i . p_i$. For simplicity, take the sample as SRS and approximate the complex statistic ‘poverty rate’ as an ordinary proportion. In case, since the design effects due to departures from SRS are likely to be very similar for the various statistics being considers, neglecting them should not substantially affect the conclusions.

Under the above assumptions, variance of the consolidate poverty rate p is given by

$$\mathrm{var}(p) = \Sigma_i W_i^2 . \mathrm{var}(p_i) + 2.\Sigma_{j<i} W_i W_j . \mathrm{cov}(p_i, p_j).$$

By considering the poverty indicator variables $p_{i,k} = \{0,1\}$ for individuals j in the population, the above equation becomes

$$\mathrm{var}(p) = \Sigma_i W_i^2 . p_i . (1 - p_i) + 2.\Sigma_{j<i} W_i W_j . p_j . (1 - p_i).$$

It is this variance that we compare with the variance of a rate (say, p_2) computed using a single poverty line such as 60% of the median, as is normally done: $\mathrm{var}(p_2) = p_2 . (1 - p_2)$. The ratio

$$g_V = (\mathrm{var}(p) / \mathrm{var}(p_2))^{1/2}$$

gives the required factor by which the standard error is reduced.

The ‘constant’ weights may come from poverty rates estimated at the country level, and then the same weights applied to each region. An appropriate choice is (Verma *et al.* 2005):

$$W_1 = \frac{1}{3}.\left(\frac{p_2}{p_1}\right),\ W_2 = \frac{1}{3},\ W_3 = \frac{1}{3}.\left(\frac{p_2}{p_3}\right)$$

where subscripts 1, 2 and 3 refer to the rates computed at the national level with poverty line thresholds, respectively, as 50, 60 and 70% of the national median equivalised income.

6.3 Reduction due to aggregation over waves for a given panel (subsample)

Of course, we cannot merely add up the sample seizes over waves in a panel survey since there is a high positive correlation between waves which reduces the gain from cumulation. Consider two adjacent waves, with proportion poor as p and p', respectively, with the following individual-level overlaps between the two waves:

	Wave w+1		
Wave w	Poor (p'_i=1)	Non-poor (p'_i=0)	total
Poor (p_i=1)	a	b	p=a+b
Non-poor (p_i=0)	c	d	1-p=c+d
total	p'=a+c	1-p'=b+d	1=a+b+c+d

Indicating by p_j and p'_j the {1,0} indicators of poverty of individual j over the two waves, we have, with the sum over all (g) individuals:

$$\mathrm{var}(p_j) = \Sigma(p_j - p)^2 / g = p.(1-p) = v_1;$$

$$\mathrm{cov}(p_j, p'_j) = \Sigma(p_j - p).(p'_j - p')/g = a - p.p' = c_1.$$

For data averaged over two adjacent years (and ignoring the difference between p and p'), variance is given by: $v_2 = \frac{1}{4}.(v_1 + v_1 + 2.c_1) = \frac{v_1}{2}.\left(1 + \frac{c_1}{v_1}\right)$. The correlation $(c_1/v_1) = R_1$ between two periods is expected to decline as the two become more widely separated. Let $(c_i/v_1) = R_i$ be the correlation between two points i waves apart. A simple and reasonable model of the attenuation with increasing i is: $(c_i/v_1) = (c_1/v_1)^i$.

Now in a set of Q periods (waves) there are (Q-i) pairs exactly i periods apart, i=1 to (Q-1). It follows from the above that variance v_Q of an average over Q periods relates to variance v_1 of the estimate from a single wave as:

$$f_c^2 = \left(\frac{v_Q}{v_1}\right) = \frac{1}{Q}.\left(1 + 2.\Sigma_{i=1}^{Q-1}\left(\frac{Q-i}{Q}\right).\left(\frac{c_1}{v_1}\right)^i\right), \text{ with } \left(\frac{c_1}{v_1}\right) \approx a - p^2,$$

where a is the overall rate of persistent poverty between pairs of adjacent waves (averaged over Q-1 pairs), and p is the (cross-sectional) poverty rate averaged over Q waves. Averaging over Q waves increases the effective sample size by $\left(1/f_c^2\right)$.

6.4 Reduction from averaging over rounds in a rotational design

Consider a rotational sample in which each unit stays in the sample for n consecutive periods, with the required estimate being the average over Q consecutive periods, such as Q=4 quarters for annual averages. The case n=1 corresponds simply to independent samples each quarter. Under the simplifying assumption of uniform variances, variance of the estimate of average over Q period is $V_a^2 = V^2/Q$.

In the general case, the total sample involved in the estimation consists of (n+Q-1) independent subsamples. These correspond to the rows in the figures below. Each subsample is 'observed' over a certain number of consecutive periods within the interval (Q) of interest.[2] In principle, for a given subsample the sample cases involved in these 'observations' are fully overlapping. The distribution of the (n+Q-1) subsamples according to the number of observation (m) provided is:

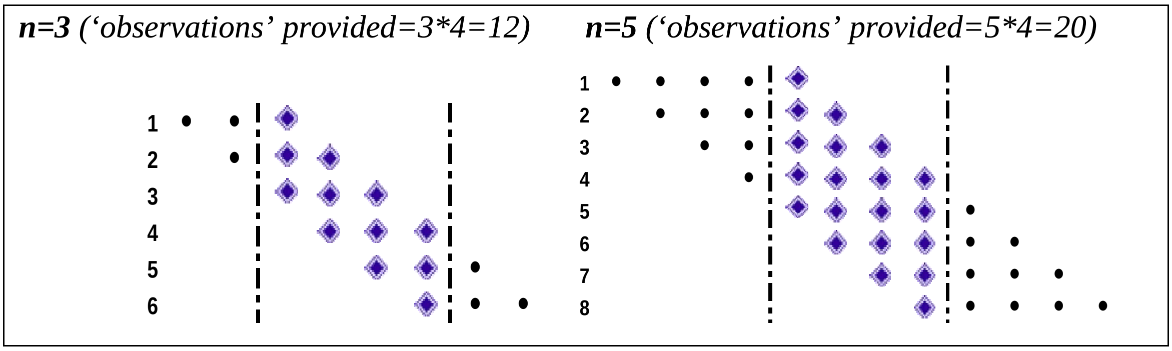
[2] For 'observation' we mean surveying one subsample on one occasion. These correspond to individual diamonds in the figures below.

No. of observations (m)→	provided by no. (x) of subsamples	Total no. of 'observations' provided by all subsamples
m = 1, 2, …, (m_1-1)	x = 2 for each value of m	$\sum_{i=1}^{(m_1-1)} 2i = (m_1-1)\cdot m_1$
m = m_1	x = m_2-(m_1-1)	$m_1 \cdot m_2 - (m_1-1)\cdot m_1$
Total →	no. of sublamples equal to $2\cdot(m_1-1)+m_2-(m_1-1)=$ $=m_2+m_1-1=n+Q-1$	no. of observations equal to $m_1 \cdot m_2 = n.Q$

where m_1=min(n, Q) and m_2=max(n, Q).

Note that the total number of 'observations' provided by all subsamples over interval Q is $m_1 \cdot m_2 = n \cdot Q$. This is consistent with the fact that, obviously, there are n subsamples observed at each of the Q periods in the interval being considered (see diagrams below).

Q=4



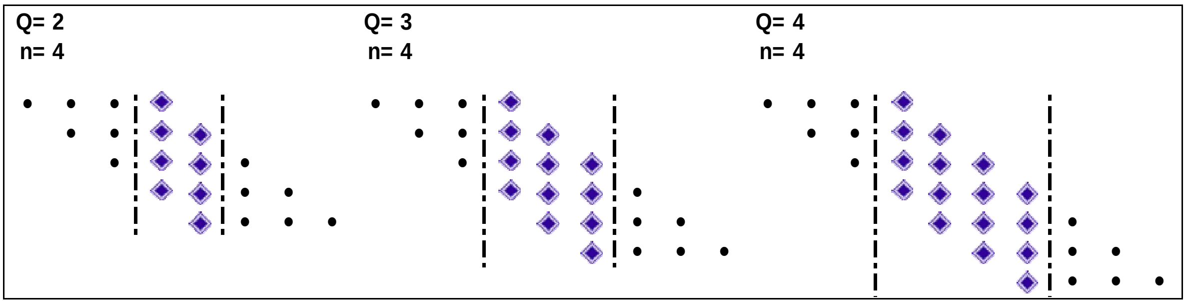
Note: The numbers on the left side of the figures represent the number of subsamples (n+Q-1).

For illustration, consider Q=m_1=4, n=m_2=5. There are 2 contributing subsamples for each number 1, 2 and (m_1-1)=3 of observations; and in addition there are m_2-(m_1-1)=2 subsamples each contributing m_1=4 observations.

Similarly, for Q=m_2=4, n=m_1=3, we have 2 contributing subsamples for each number 1 and (m_1-1)=2 of observations, and in addition m_2-(m_1-1)=2 subsamples each contributing m_1=3 observations.

In the EU-SILC survey in most countries, n is always equal to 4 (each survey rounds is made of 4 subsamples), and at the present stage Q could be equal to 2 (years 2003-2004), 3 (years 2003-2004-2005) and 4 (years 2003-2004-2005-2006).

So the previous figure could be adapted as follow:



In order to provide a simplified formulation of the effect of correlation arising from sample overlaps, we assume the following model. If R is the average correlation between estimates from overlapping samples in adjacent periods (as defined above), then between points one period apart (e.g. between the 1st and 3rd quarters), the average correlations is reduced to R^2, the correlation between points two periods apart (e.g. the 1st and the 4th quarters) is reduced to R^3, and so on.

Consider a subsample contributing m observations during the interval (Q) of interest with full sample overlap. Considering all the pairs of observations involved and the correlations between them under the model assumed above, variance of the average over the m observations is given by

$$V_m^2 = \frac{V^2}{m} \cdot (1 + f(m))$$

where

$$f(m) = \frac{2}{m} \cdot \left\{ (m-1) \cdot R + (m-2) \cdot R^2 + \ldots + R^{m-1} \right\}$$

The term $V_m^2 \Big/ \left(\frac{V^2}{m} \right) = 1 + f(m)$ reflects the loss in efficiency in cumulation or averaging over overlapping samples, compared to cumulation over entirely independent samples. The following illustrates its values for various values of m:

m	*f(m)*
2	R
3	$\frac{2}{3}(2R+R^2)$
4	$\frac{2}{4}(3R+2R^2+R^3)$
5	$\frac{2}{5}(4R+3R^2+2R^3+R)$

Repeated observations over the same sample are less efficient in the presence of positive correlations (R). The loss depends on the number of repetitions (m) and is summarised by the factor *[1+f(m)]*.

In estimating the average using the whole available sample of $(n \cdot Q)$ subsample observations[3], we may simply give each observation the same weight. Taking into account the number of observations and the variances involved, the resulting variance of the average becomes:

$$V_a^2 = \left(\frac{V^2}{n \cdot Q}\right) \cdot \left\{ m_1 \cdot [m_2 - (m_1 - 1)] \cdot [1 + f(m_1)] + 2\sum_{m=1}^{m_1 - 1} m \cdot [1 + f(m)] \right\} \Big/ (n \cdot Q) = \left(\frac{V^2}{n \cdot Q}\right) \cdot F(R)$$

The first factor is the variance to be expected from $(n \cdot Q)$ independent observations (with no sample overlaps or correlation), each observation with variance V^2. The other terms are the effect of correlation with sample overlaps. This effect, *F(R)* disappears when *f(i)=0* for all *i=1 to m* (which will be the case of *R=0*), as can be verified in the above expression.

An alternative is to take a weighted average of the observations, with weights inversely proportional to their variance, i.e. to the corresponding factor *[1+f(m)]*. The effect on the resulting variance, though may appear algebraically cumbersome, can be easily worked out, for any given rotation pattern and value of average correlation R.

[3] Obviously , we have n subsamples observed during each of Q periods in the rotational design assumed.

It has the form

$V_a^2 = \sum W_i^2 \cdot V_i^2$, with $\sum W_i = 1$, where W_i are the relative weights given to observations in a set involving I repetitions during the interval of interest.

7. Concluding remark: objectives of pooling

It may be argued that averaging and similar 'manipulation' is not acceptable, or at least that it introduces bias, since it *alters* the measures we obtain. This may be true in a literal sense but this is not a sensible objection in many situations. We need a pragmatic and not an ideological approach to statistics. All statistical measures are constructed for the purpose of conveying some meaning, for providing some interpretation to real and complex situations. The particular forms of measures chosen are always determined by considerations of usefulness and practicality, are always compromises and in themselves not 'sacred'. The objectives of pooling include searching for measures which convey essentially the same information as the 'original' un-pooled measures, but in a *more robust* manner, reducing random variability or noise. A related objective of pooling is *trading dimensions* – gaining in some more needed directions by losing something less needed for the particular purpose – such as permitting more detailed geographical breakdown but with less temporal detail. A third objective is to *summarise* over different dimensions, providing more consolidated and fewer indicators. Such indicators are of course different from the more numerous 'raw' indicators, but are often more, or at least equally, meaningful and useful.

References

Verma V., (1999), Combining national surveys for the European Union, *Proceedings of the 52nd Session of the International Statistical Institute*, Helsinki.

Verma, V. (2002), Comparability in Multi-country Survey Programmes. *Journal of Statistical Planning and Inference*, **102(1)**, pp. 189-210.

Verma V. (2005), *Indicators to reflect social exclusion and poverty* Report prepared for Employment and Social Affairs DG - with contribution of Gianni Betti, Achille Lemmi, Anna Mulas, Michela Natilli, Laura Neri and Nicola Salvati.

Verma V. and Betti G. (2006), EU statistics on income and living conditions (EU-SILC): choosing survey structure and sample design. *Statistics in Transition*, **7(5)**, pp. 935-970.

Verma V., Gagliardi F. and Ciampalini G. (2009), Methodology of European labour force surveys: (3) Sample rotation patterns. *DMQ Working Paper* 80/09, University of Siena.