Task force on quality of BCS data

Analysis of sample size in consumer surveys

theoretical considerations and factors determining minimum necessary sample sizes, link between country size and sample size (impact of larger structural heterogeneity), gross versus effective sample size (i.e. correcting for response rates), empirical evidence on links with data volatility (and possibly bias)

August 2013

Krzysztof Puszczak
Agnieszka Fronczyk
Marek Urbański

Table of Contents

1. Introduction

1.a Theoretical aspects

A common aim of survey research is to obtain data representative of a population. Researchers use information collected from a survey to generalize results from a sample to a population within the limits of random error. Designing a survey is a demanding task as a survey itself is a complex process. Apart from defining a research problem, selection of methodology and developing a questionnaire, defining a sample is the fundamental stage. In common practice the examined sample always generates particular costs due to its accessibility and size, i.e. the number of respondents who will participate in the survey. The sample size is chosen to increase the chance of uncovering a definite mean difference, which could also be statistically significant. The reason why a larger sample increases a chance of significance is because it more reliably reflects a population mean.

The answer to the problem - how to determine the size of a sample - depends on a number of factors, including the purpose of the study, population size, the risk of selecting a wrong sample, and the acceptable sampling error. In addition to population size and the goal of the study, usually three criteria are necessary to be specified in order to determine the sample size:

- sampling error
- the confidence level
- and the degree of variability in the main measured attributes

Sampling error (or the level of precision) is the range in which the real value for the population is estimated. Usually this range is expressed in percentage points. Thus, if a result of a study shows that 35% of sample plan to spend money on education it means that between 31% and 39% of the population have such plans.

The confidence level is based on ideas of the Central Limit Theorem (CLT). The key idea encompassed in the CLT is that when a population is repeatedly sampled, the average value of the attribute obtained by those samples is equal to the real value for the population. What's more, the values obtained by these samples are distributed normally around the real value. In a normal distribution, approximately 95% of the sample values are within two standard deviations of the real value for the population (chart. 1.0).

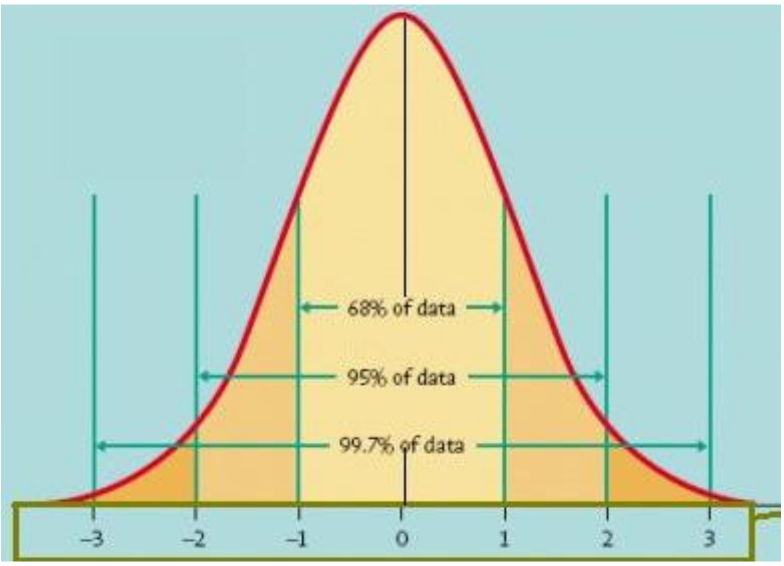

Chart 1.0 Distribution of means (x=0) for repeated samples, unit=standardized deviation

.

The degree of variability in the measured attributes, applies to the distribution of attributes in the population. The more heterogeneous a population, the larger the sample size required to reach an acceptable sampling error. The more homogeneous a population, the smaller the sample size. As the proportion of 50% indicates the maximum variability in a population, it is often used in determining a more conservative sample size. Therefore the sample size may be larger than if the true variability of the population attribute were used.

Aiming to determine an exact samples size, some possible ways could be chosen:

- using or conducting **a census survey** for small populations - the entire population makes up the sample (for small populations, e.g., 200 or less)
- a sample size transferred from **a similar study** – using a same sample size as those of studies similar to the planned one
- relying on **published tables**, which provide the sample size for a given combinations of sampling error, confidence levels and variability
- applying of **formulas to calculate a sample size** for a different combination of sampling error, confidence and variability.

Generally researchers know from their experience, that first, two ways are feasible only under some particular conditions: a small population or a previous survey on similar subject, respectively. Much more universal are two other solutions. Specifying sampling error, confidence levels and variability one can easily find the correct sample size basing on published tables or formulas. Using formulas can be also helpful in checking how particular size of a sample will affect estimations in case when sample size is initially determined from any reasons. The tables are widely accessible and easy to apply. Published tables themselves are based on specific formulas which can be used directly to determine a sample size. Therefore this often preferred in researchers daily practice. Below there some formulas very helpful in obtaining desired sample size.

For populations that are large, Cochran (1963:75) developed the Equation 1.1 to yield a representative sample for proportions.

$$n_0 = \frac{Z^2 * p * (1 - p)}{e^2}$$

Equation 1.1 Formula for calculating a sample for proportions

n_0 - the sample size
Z^2 - the abscissa of the normal curve that cuts off an area α at the tails
e - the acceptable sampling error
p - the estimated proportion of an attribute that is present in the population

If the population is small then the sample size can be reduced slightly. This is because a given sample size provides proportionately more information for a small population than for a large population. The sample size (n_0) can be adjusted using Equation 1.2.

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

Equation 1.2 Finite population correction for proportions

n - adjusted sample size
N - the population size.

In common practice probably the most popular is another equation from late 60s. Yamane (1967:886) provides a simplified formula to calculate sample sizes. A 95% confidence level and P = .5 are assumed for Equation 1.3.

$$n = \frac{N}{1 + N * (e)^2}$$

Equation 1.3 A simplified formula for proportions

n - the sample size
N - the population size
e - the acceptable sampling error

Sometimes examined attributes are measured on interval scale where mean represents a central tendency for an attribute. There is a need to apply another variant of the equation 1.1, then. This formula of the sample size for the mean is similar to that of the proportion, except for the measure of variability. The formula for the mean employs σ^2 instead of (p x q), as shown in equation 1.4.

$$n_0 = \frac{Z^2 * \sigma^2}{e^2}$$

Equation 1.4 Formula for sample size for the mean

n0 - the sample size
Z - the abscissa of the normal curve that cuts off an area σ at the tails
e - the acceptable sampling error
σ2 - the variance of an attribute in the population.

The disadvantage of the sample size based on the mean is that a "good" estimate of the population variance is necessary. Often, an estimate is not available. Furthermore, the sample size can vary widely from one attribute to another because each is likely to have a different variance. Because of these problems, the sample size for the proportion is frequently preferred.

On the basis of the theory determining a sample size is not very complex. However some additional problems appear. They are mainly caused by some factors dependent on a survey's specificity. Therefore is very frequent when theoretical approach is enough. For instance, formulas from above (equations 1.1 1.3. 1.4) are valid under assumption of simple random sample as the sampling design. In case of more complex designs, for instance stratified random samples, the variances of subpopulations, strata, or clusters have to considered to estimate of the variability in the population. In case of need of some subgroups comparison an adjustment of a sample size could be indispensable. In the literature one can find some hints about this problem. For instance Sudman (1976) suggests that a minimum of 100 observations is necessary for a major group or subgroup in the sample and for a minor subgroup, a sample of 20 to 50 observations is desired. Similarly, Kish (1965) suggest that 30 to 200 elements are sufficient when the attribute is present 20 to 80 percent of the time (i.e., the distribution approaches normality). Skewed distributions (far from normality) can also cause need of increase in size samples (Kish, 1965).Formulas to calculate a sample size provide the number of responses that need to be collected. Usually it is correct to increase a sample size to compensate lack of interviews caused by inaccessibility. The sample size should be also increased to compensate for nonresponse. Therefore it is necessary enlarge calculated sample size to reach a desired level of confidence with assumed sampling error.

In general practice, the larger the sample size the higher precision is obtained to make conclusions about a population. However the improvement in accuracy for larger sample sizes is minimal, almost negligible. The illustrative example of this relation is the chart showing dependence of sampling error on sample size.

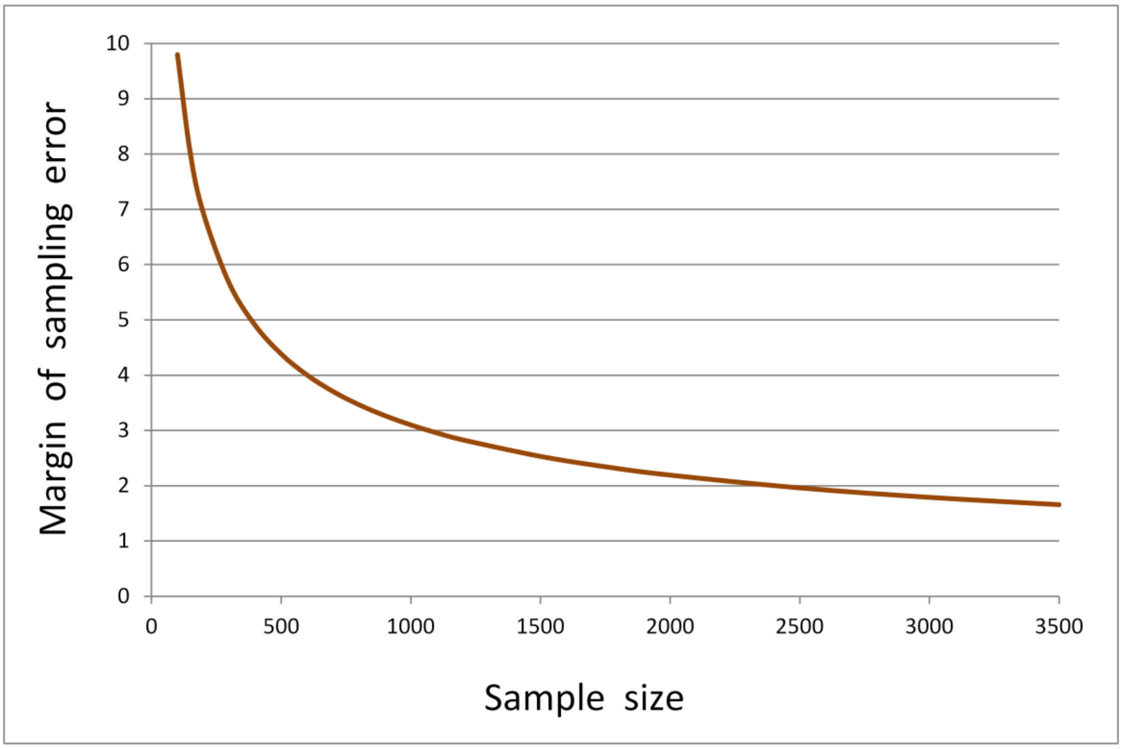

Chart 1.1 dependence of sampling error on sample size. Margin of sampling error was calculated for 95% confidence interval, 50% fraction and infinite population.

It is clearly visible that gaining more accuracy is relatively slower when sample size increases.
Generally in social studies an appropriate sample size is a key factor that determines the accuracy of study results.
However the gains in precision of estimates are not directly proportional to increases in sample size (doubling the sample size will not halve the standard error (see Equation 1.1), generally the sample has to be increased by a factor of 4 to halve the SE). In practice the most relevant factor which limits a sample size is the cost of the study or to be more exact: marginal cost of precision.

1.b Sample size in the DG ECFIN Consumer Survey

In this section the range of sample sizes used by the institutes participated in the DG ECFIN Consumer Survey are presented. All the results are based on the excel file (*Metadata_checked_by_partners - Consumers.xlsx*) that contains information about the sample design delivered by the institutes. Due to some inaccuracy in this file the presented results should be treated as tentative and need to be supplemented after verification of the source data.

Following given declarations, the smallest sample size is 500 and the largest 3300. All the values are shown in the chart below:

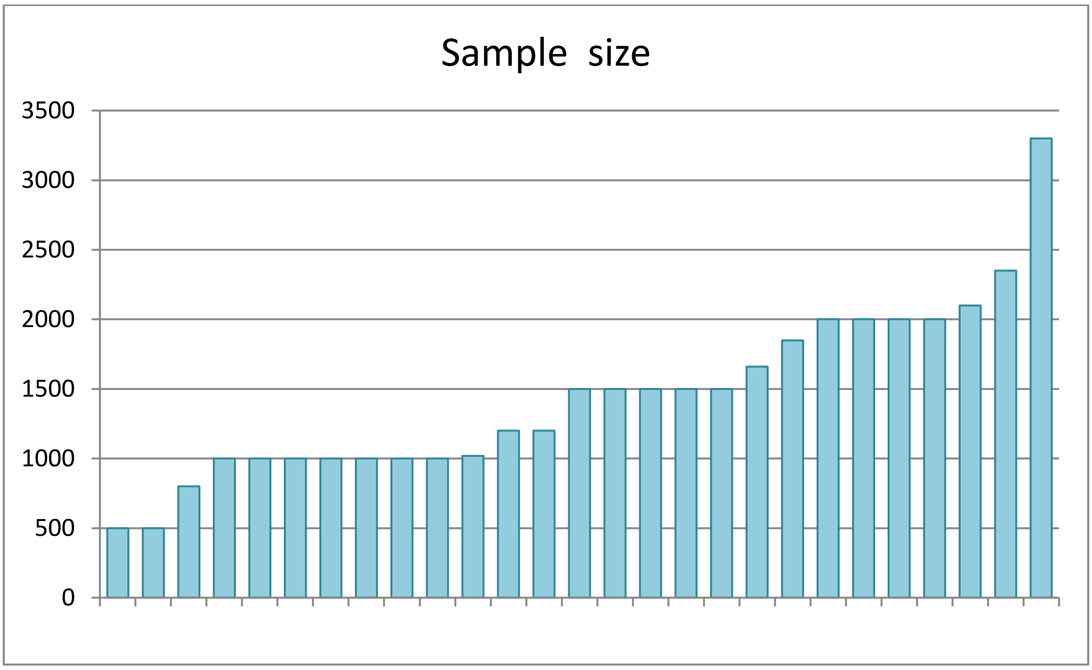

Chart 1.2 Distribution of sample size, n=27.

Other information with key importance is effective sample size (ESS). Presented numbers are likely not be delivered directly from the institutes but determined with one algorithm. According to them the smallest ESS is 118 and the highest is 2310. Nevertheless both distribution of sample size and ESS across countries looks promising regarding research on relations with quality measures, i.e. volatility of the data and the tracking performance with respect to statistical reference series (Chart 1.3).

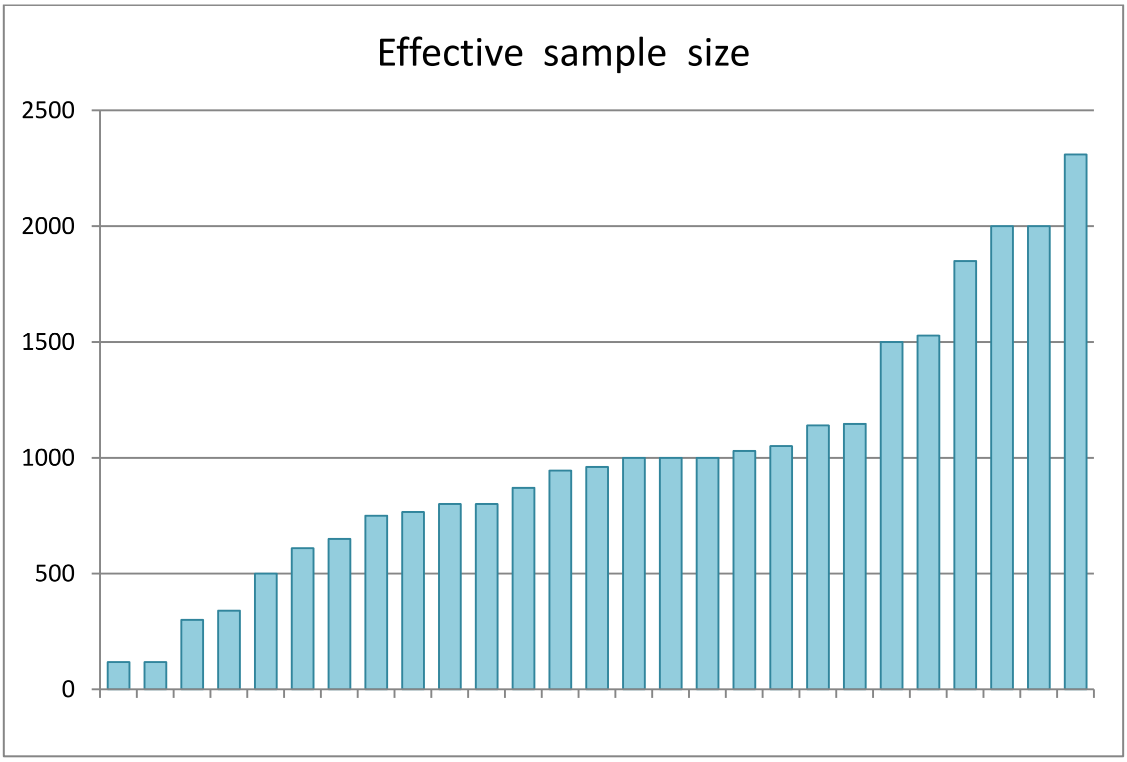

Chart 1.3 Distribution of effective sample size, n=27.

2. Quality measures vs. sample size

In this section the influence of sample size used by the institutes participated in the DG ECFIN Consumer Survey on quality measures is examined. Looking into the distributions of quality measures, i.e. volatility of the data (measured with MCD 1 and MCD 2) and the tracking performance with respect to statistical reference series (measured with correlation to these series) some fairly considerable variability is visible (Charts 2.1 – 2.3). This observation lets us assume that there are potential drivers of quality measures in the analysed data.

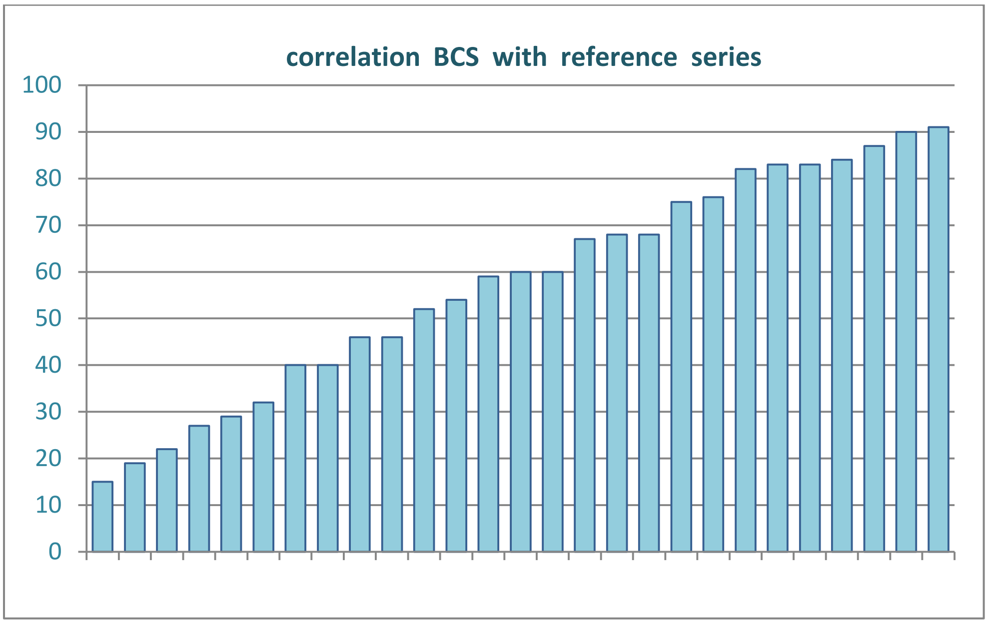

Chart 2.1. distribution of correlation BCS with reference series, n=27

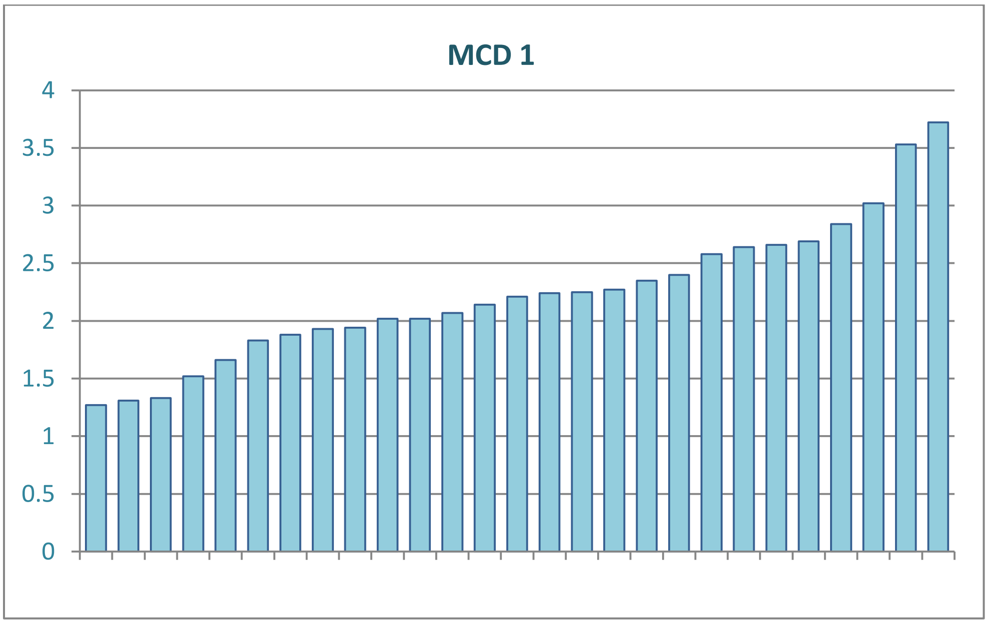

Chart 2.2. distribution of MCD 1, n=27

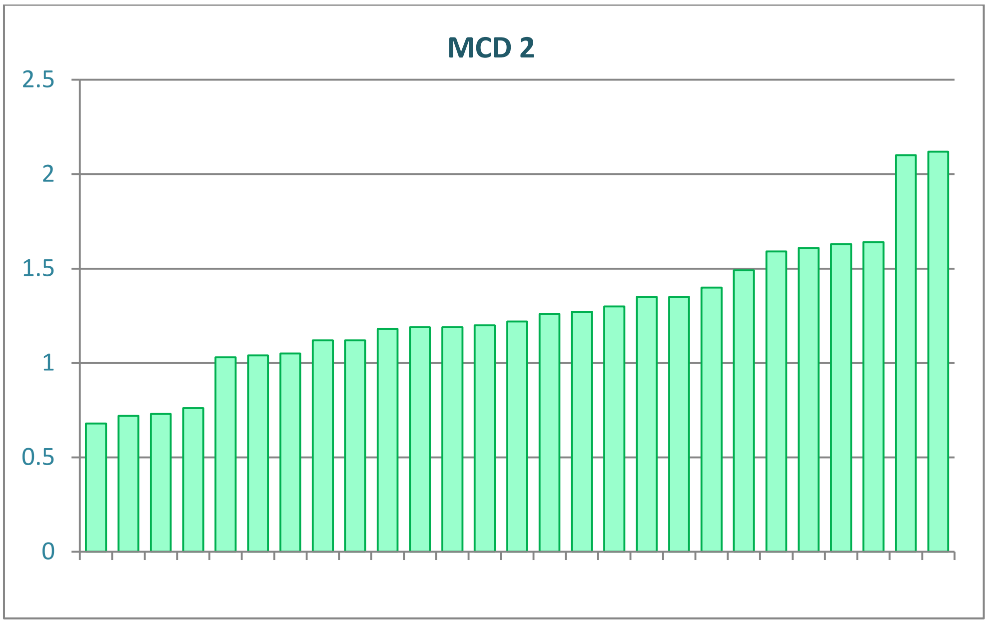

Chart 2.3. distribution of MCD 2, n=27

In the next paragraphs some results of analyses based on relations between quality measures and sample size or ESS will be presented.

2.a Sample size - analysis of impact

In the framework of diagnostic stage the basic scatter plots were prepared. This first view reflects the real data from one side and gives an opportunity to discover a general type of possible relations between two variables on the other side.

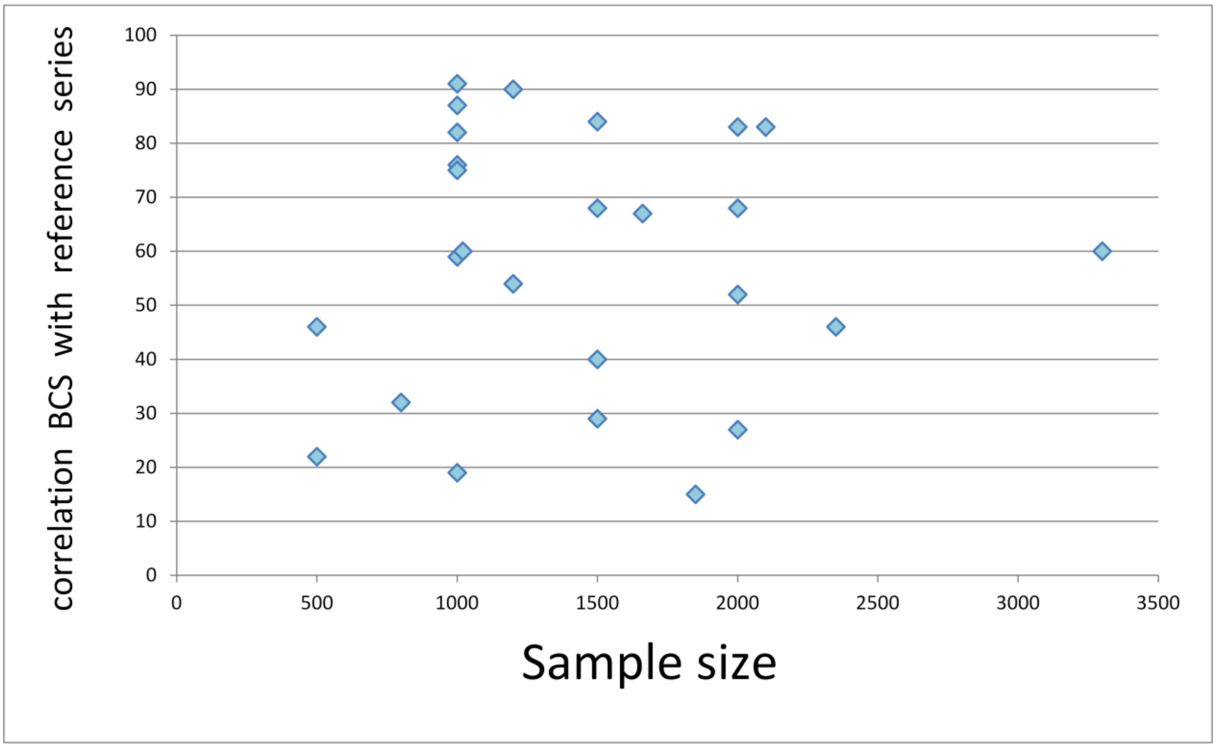

Chart 2.3. .scatter plot for sample size and correlation BCS with reference series, n=27

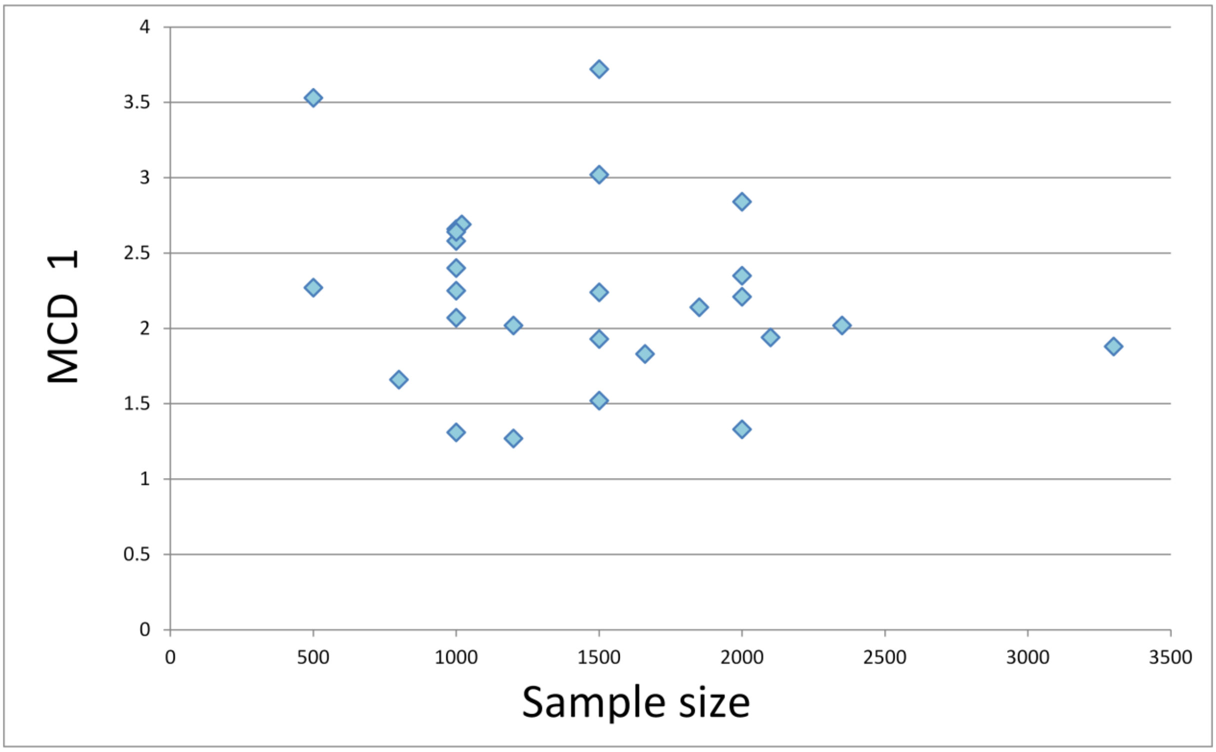

Chart 2.4. .scatter plot for sample size and MCD 1, n=27

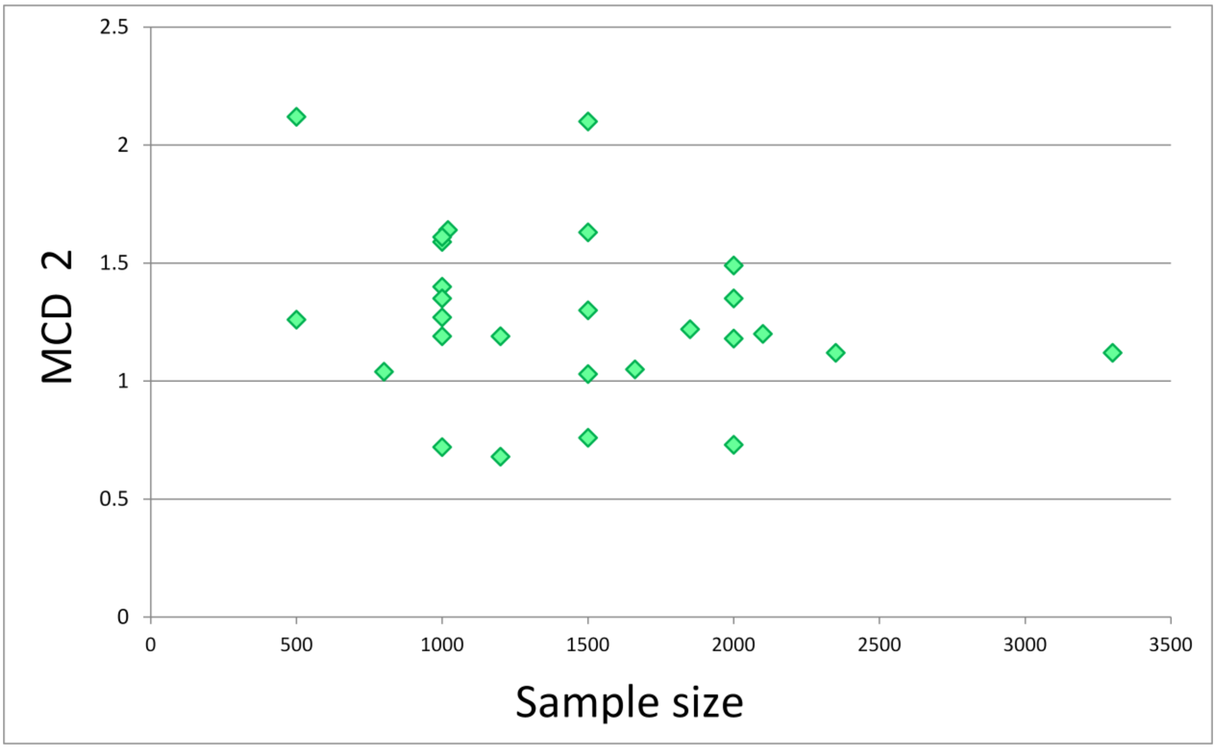

Chart 2.5. .scatter plot for sample size and MCD 2, n=27

For all quality measures there is no clear pattern of their connection with sample size. According to the charts 2.3 to 2.5 it would be risky to prove that larger sample size causes better quality of data. Therefore another perspective of presenting these values is proposed.

Starting from ranking all the samples from the smallest size to the highest one the possibility to show average quality measures for groups of countries is easy to produce. To make a potential tendency more stable groups of samples are cumulated i.e. each group includes all samples with smaller sizes, up to four sizes together (Table 1.1).

sample size	correlation BCS with ref.series (moving average)	MCD_1 (moving average)	MCD_2 (moving average)
500	34.0	2.90	1.69
800, 500	33.3	2.49	1.47
1000, 800, 500	58.9	2.34	1.36
1020, 1000, 800	59.0	2.37	1.38
1200, 1020, 1000, 800	65.9	2.14	1.24
1500, 1200, 1020, 1000	63.6	2.29	1.30
1661, 1500, 1200, 1020	59.1	2.25	1.26
1850, 1661, 1500, 1200	54.1	2.19	1.22
2000, 1850, 1661, 1500	52.1	2.28	1.26
2100, 2000, 1850, 1661	56.4	2.09	1.17
2350, 2100, 2000, 1850	53.4	2.12	1.18
3300, 2350, 2100, 2000	59.9	2.08	1.17

Table 1.1 quality measures as moving averages for cumulated countries' samples

This approach could answer the question whether there is a tendency expressing positive changes in the quality of data along with the increase of sample size. Such analysis brings results shown on charts 2.6 and 2.7.

Chart 2.6. X-axis: grouped samples with at most "x" sample size. Y-axis: means of correlation BCS with reference series for grouped samples, n=27

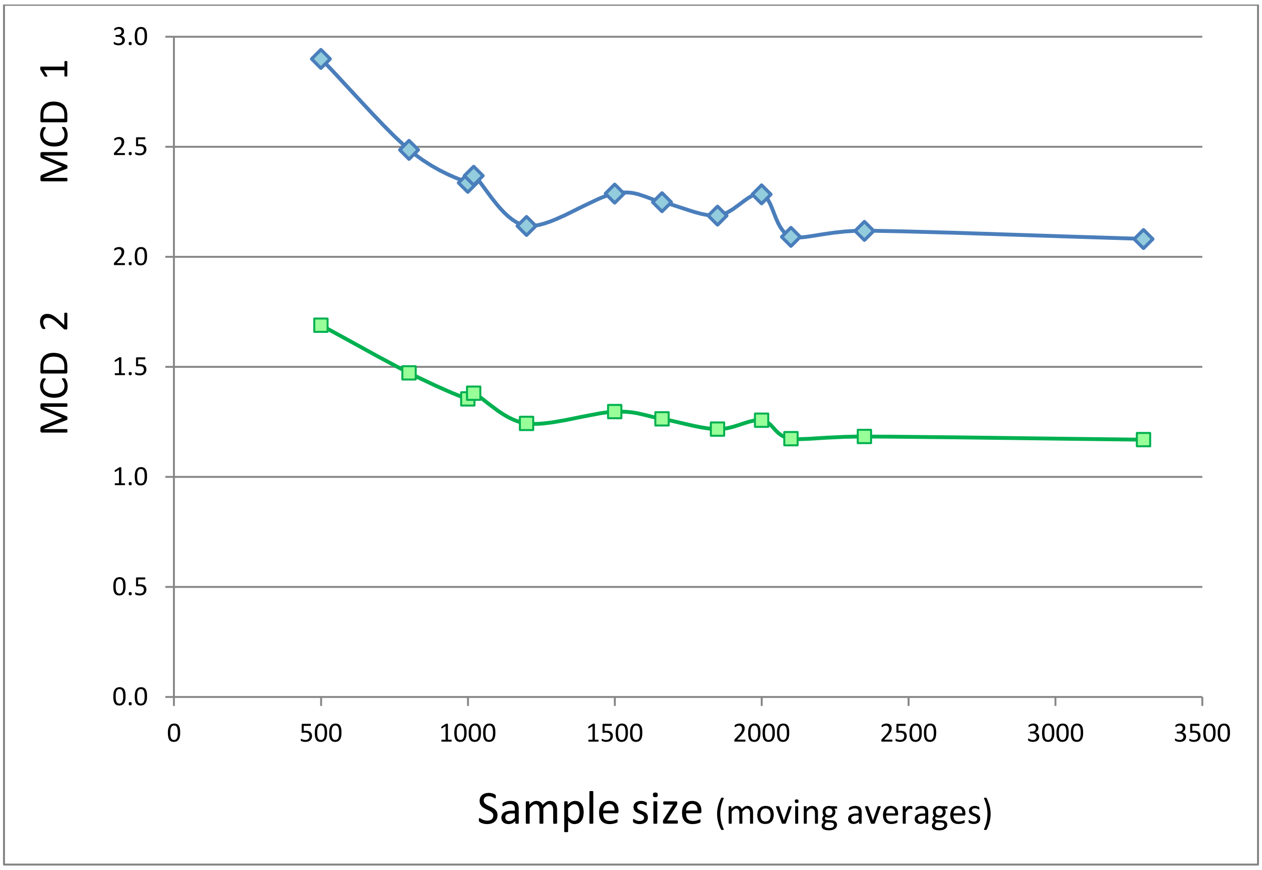

Chart 2.7 X-axis: grouped samples with at most “x” sample size. Y-axis: means of MCD_1 (upper line), MCD_2 (lower line), n=27

Presented approach seems to be more helpful in reaching conclusion. Hence, it is much easier to accept the hypothesis that the larger the sample size the better average quality measures can be expected. Such finding supports common intuition. The trend line is rather clear but this relation is not very strong. Further increase tends to bring some improvement in data quality. As a general rule it could be easy to adopt.

2.b Effective sample size - analysis of impact

Basic scatter plots showing effective sample size and quality measure is a starting point in this part. As for sample size. This view reflects the real data from one side and gives a chance to discover a general type of possible relations on the other side.

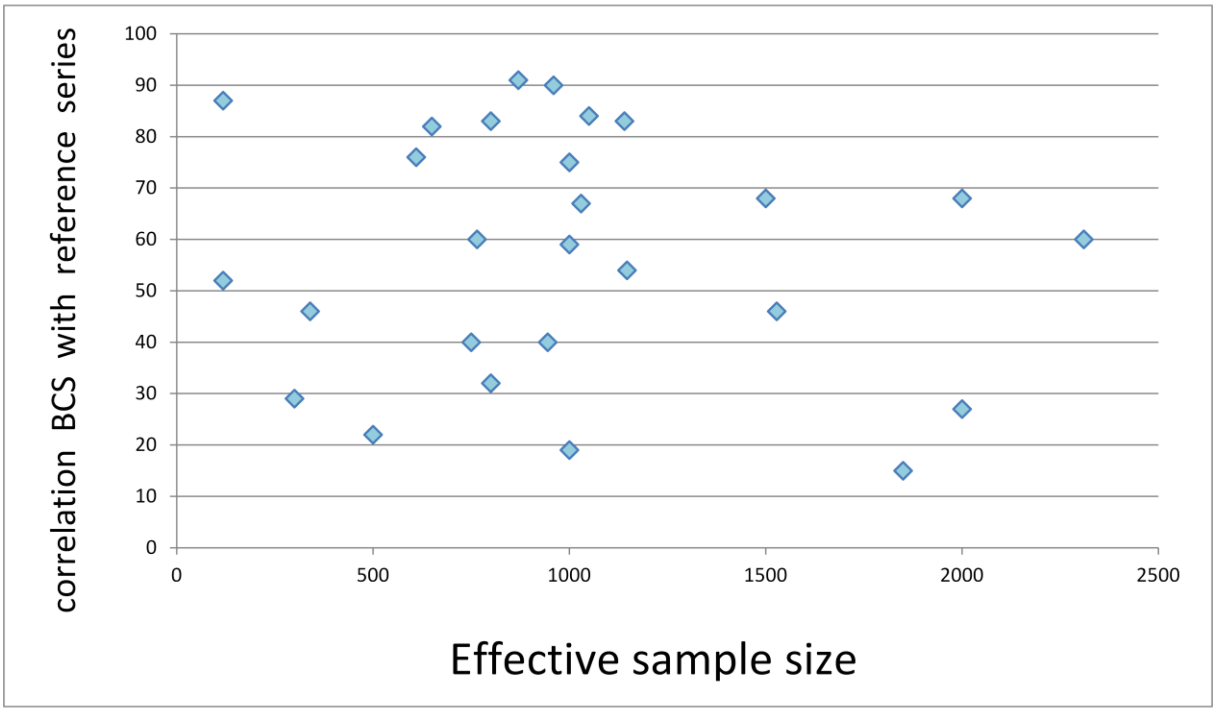

Chart 2.8. .scatter plot for effective sample size and correlation BCS with reference series, n=27

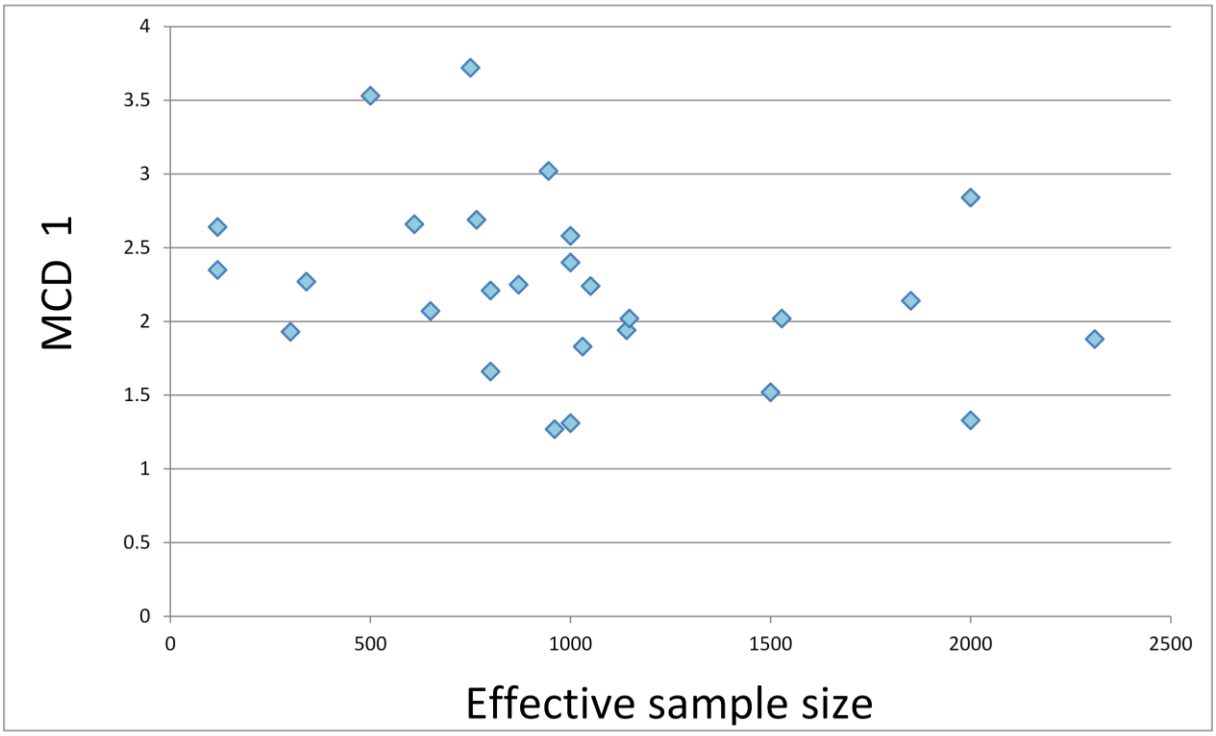

Chart 2.9. .scatter plot for effective sample size and MCD 1, n=27

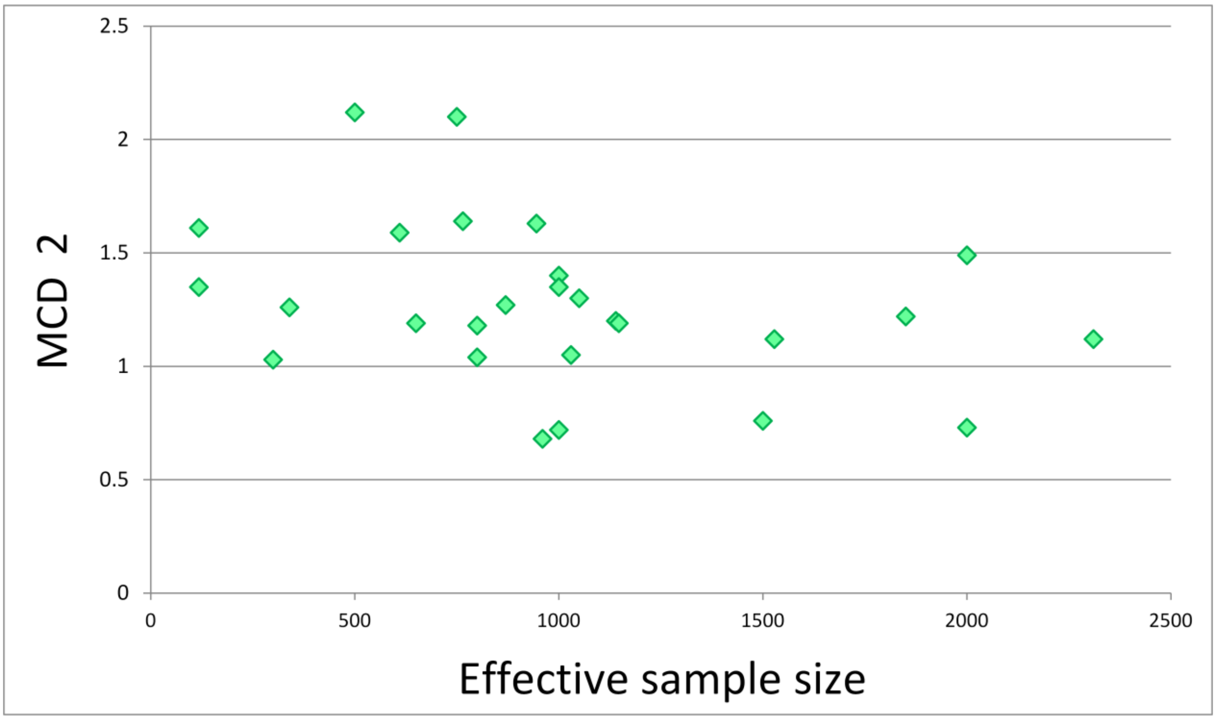

Chart 2.10. .scatter plot for effective sample size and MCD 2, n=27

As it could be expected after sample size analysis for all quality measures there is no clear pattern of their connection with effective sample size. According to the charts 2.8 to 2.10 it is not possible to prove that larger effective sample size causes better quality of data. Therefore analogical perspective of presenting these values was prepared.

sample size	**correlation BCS with ref.series (moving average)**	**MCD_1 (moving average)**	**MCD_2 (moving average)**
118	69.5	2.50	1.48
300, 118	56.0	2.31	1.33
340, 300, 118	53.5	2.30	1.31
500, 340, 300, 118	47.2	2.54	1.47
610, 500, 340, 300, 118	52.0	2.56	1.49
650, 610, 500, 340, 300, 118	56.3	2.49	1.45
750, 650, 610, 500, 340, 300	49.2	2.70	1.55
765, 750, 650, 610, 500, 340	54.3	2.82	1.65
800, 765, 750, 650, 610, 500	56.4	2.65	1.55
870, 800, 765, 750, 650, 610	66.3	2.47	1.43
945, 870, 800, 765, 750, 650	61.1	2.52	1.44
960, 945, 870, 800, 765, 750	62.3	2.40	1.36
1000, 960, 945, 870, 800, 765	61.0	2.15	1.21
1030, 1000, 960, 945, 870, 800	61.8	2.06	1.15
1050, 1030, 1000, 960, 945, 870	65.6	2.11	1.18
1140, 1050, 1030, 1000, 960, 945	64.6	2.07	1.17
1147, 1140, 1050, 1030, 1000, 960	66.4	1.95	1.11
1500, 1147, 1140, 1050, 1030, 1000	63.6	1.98	1.12
1528, 1500, 1147, 1140, 1050, 1030	67.0	1.93	1.10
1850, 1528, 1500, 1147, 1140, 1050	58.3	1.98	1.13
2000, 1850, 1528, 1500, 1147, 1140	51.6	1.97	1.10
2310, 2000, 1850, 1528, 1500, 1147	48.3	1.96	1.09

Table 1.2 quality measures as moving averages for cumulated countries' ESS

The charts below are based on ranking all samples from the smallest ESS to the highest. Samples were grouped to present average quality measures for them (Table 1.2). Expecting any tendency groups of samples are cumulated i.e. include all samples with smaller ESS, up to six sizes together. It can be possible to observe if positive changes in quality of data are accompanied by the increase of ESS. Such analysis brings results shown on charts 2.11 and 2.12.

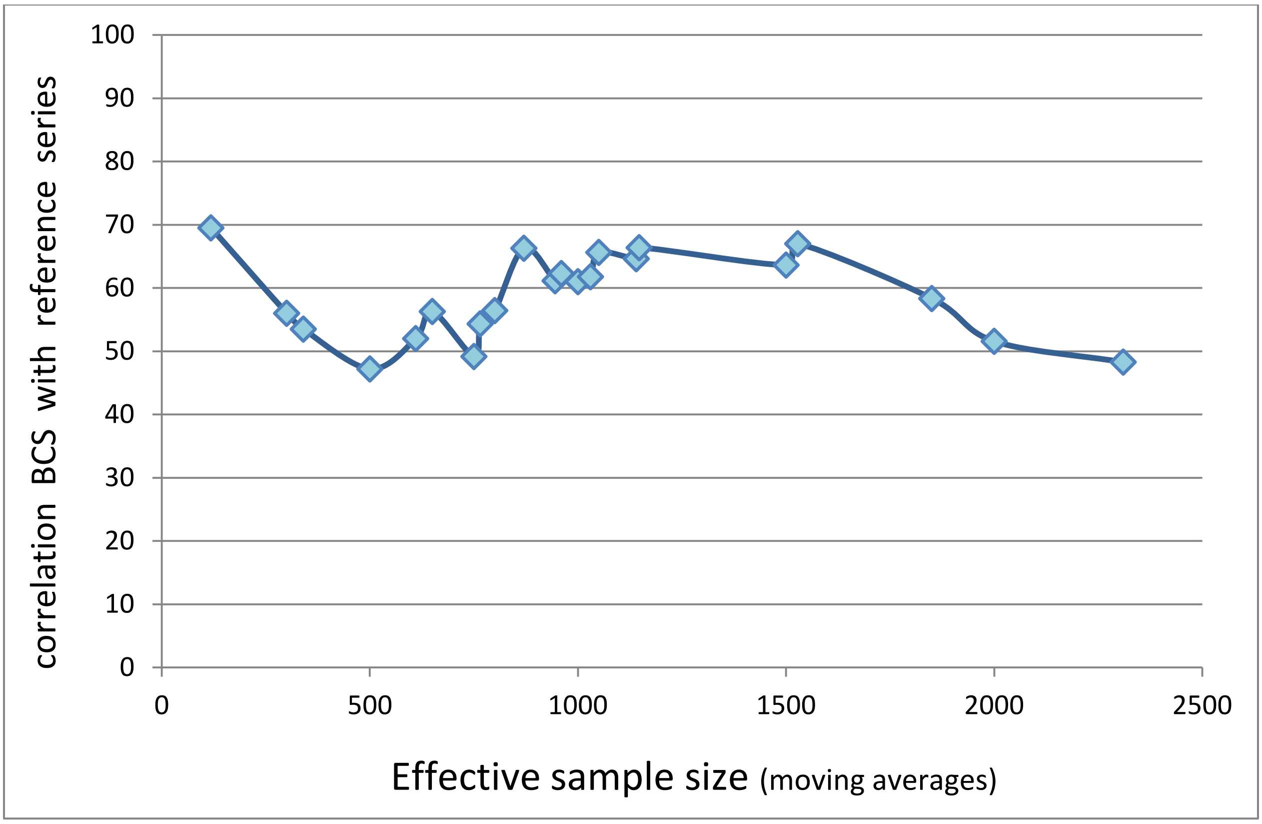

Chart 2.11. X-axis: grouped samples with at most “x” effective sample size. Y-axis: means of correlation BCS with reference series for grouped samples, n=27

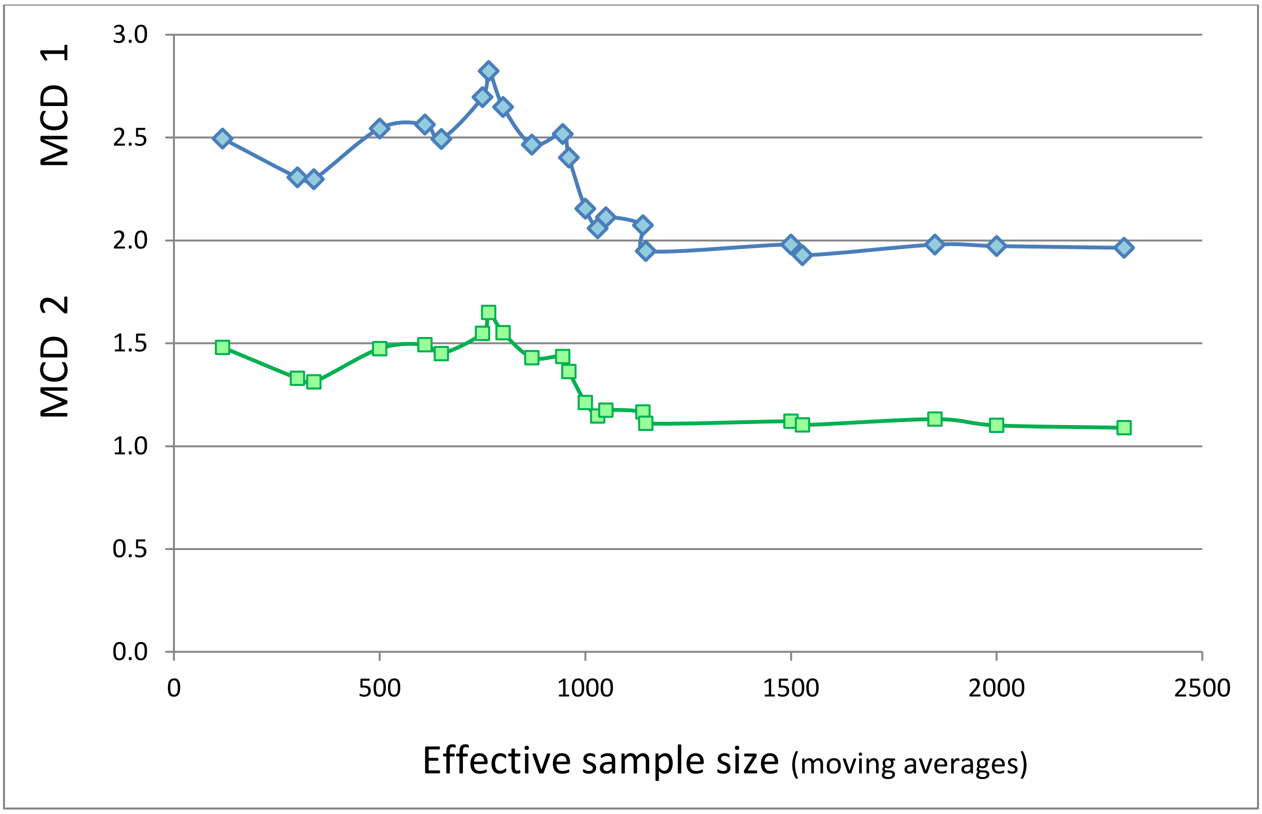

Chart 2.12 X-axis: grouped samples with at most “x” effective sample size. Y-axis: means of MCD_1 (upper line), MCD_2 (lower line), n=27

The proposed perspective again turned out to be rather helpful in reflecting the overall tendency. The possible connection between ESS and quality measures is not so clear as in case of sample size. The major disturbance is caused by cases of the smallest ESS. Both cases of ESS equal 118 have on average very high quality measures (the one close to total average, the other much higher). This is the reason why the tendency on the whole range is not very obvious. Nevertheless, for both MCD measures, starting the interpretation from ESS labelled *765* (with the highest MCD level) within ESSs of 765, 750, 650, 610, 500 and 340, the expected relations are quite visible. As before it is still not very strong linkage or alternatively a hypothetical causality. Considering relation of ESS with correlation BCS with reference series, the possible connection between ESS and quality measures is not as clear as in case of sample size. At least it is an acceptable effect of coincidence the larger effective sample size and the better average quality measures (MCD mainly). Such observation supports the previous finding of sample size effect. As before this moderate relation becomes much stronger as tendency, especially on the great part of the range. There is not definitive arguments for statement that increase of ESS causes improvement in data quality. However this is more probable conclusion than opposite one.

2.c Additional analysis

In order to go deeper into the data one additional verification was conducted. It was based on hypothesis that the potential impact of ESS on quality measures varies in different sample frames.

Based on the given description of sampling frames the following types of sampling frames (SF) were defined:

- telephones access as unit of SF, represented by 15 cases
- individuals as unit of SF, represented by 10 cases
- SF not classified in 6 cases

Regarding missing information about quality measures for four countries the numbers for the types are respectively: 14, 9 , 4 .

Prepared three scatter plots are expected to shed some light on the mentioned above different relations.

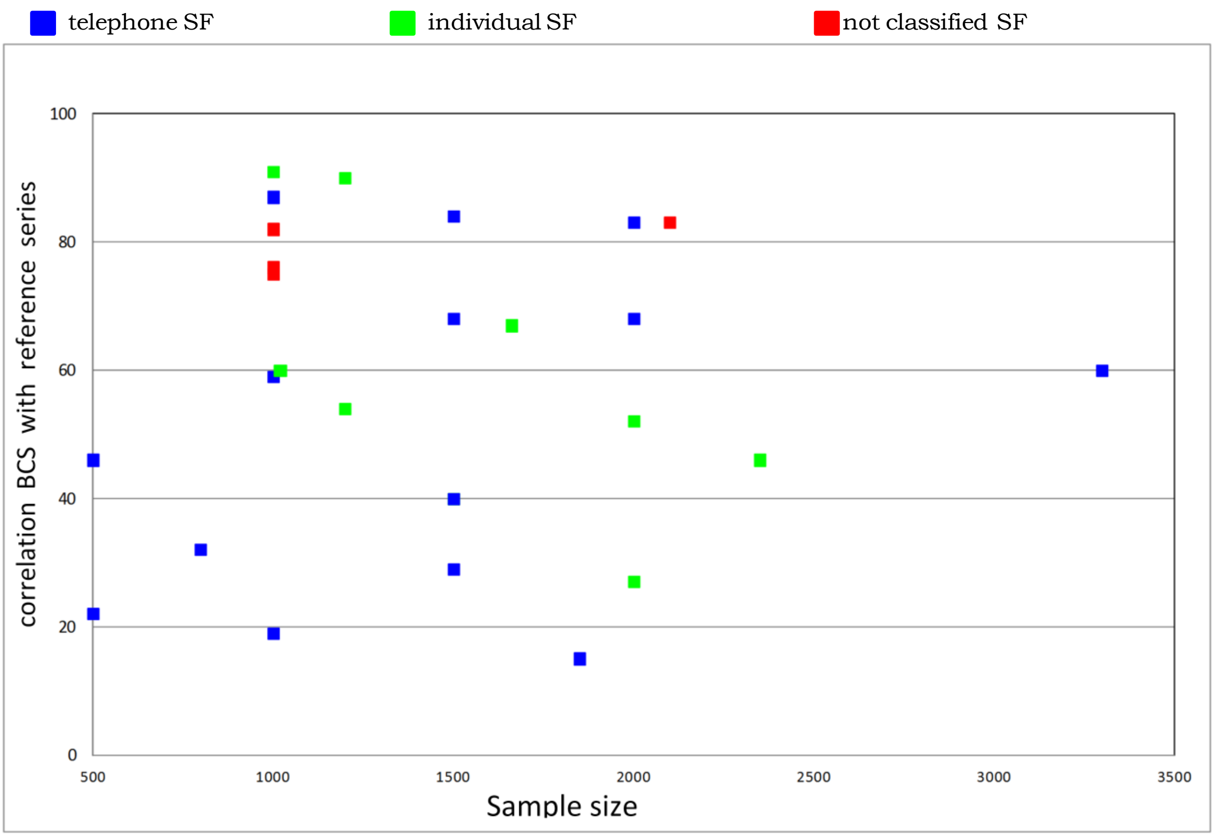

Chart 2.13 .scatter plot for sample size and correlation BCS with reference series for 3 types of sampling frames, n=27

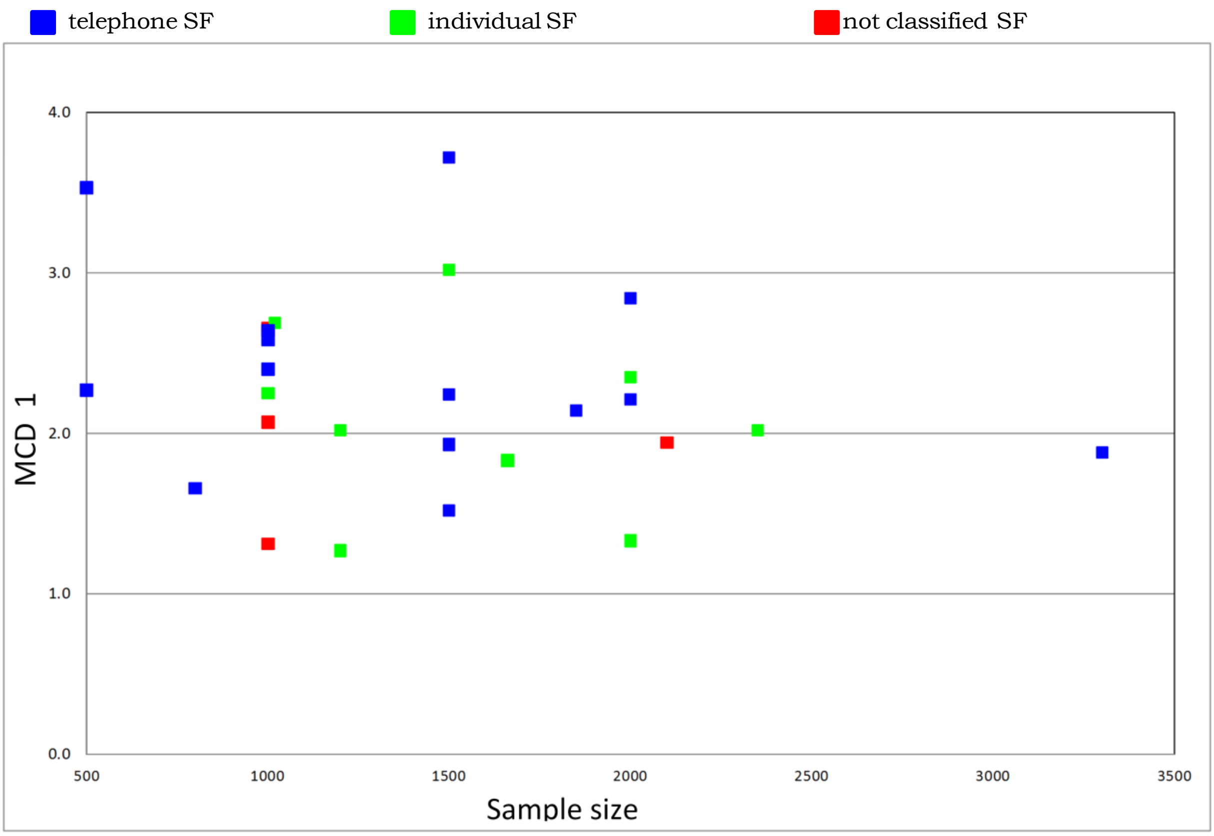

Chart 2.14 .scatter plot for sample size and MCD_1 for 3 types of sampling frames, n=27

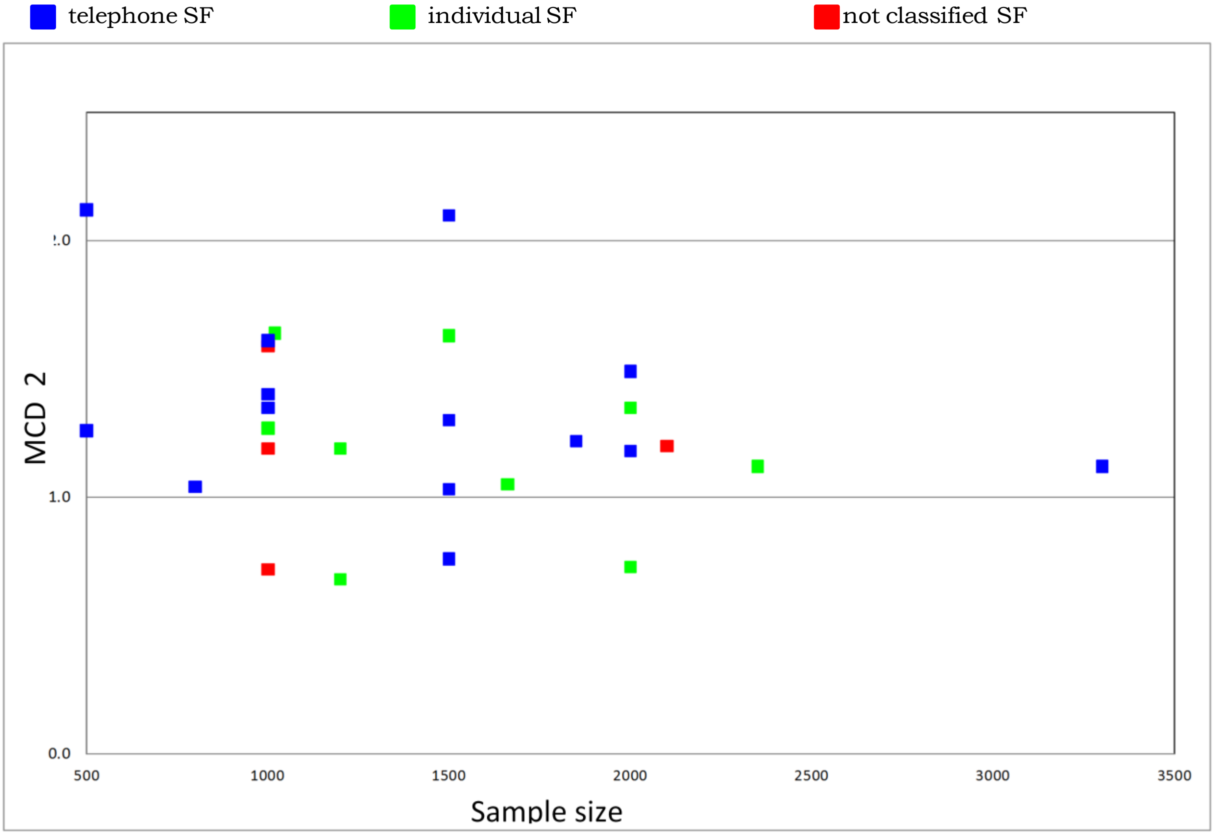

Chart 2.15 .scatter plot for sample size and MCD_2 for 3 types of sampling frames, n=27

This additional approach did not contribute significantly to the conducted deliberation. What's more on the chart 2.13 the odd pattern is observed for sampling frames with individuals. A questionable, although quite clear, relation shows that the larger the sample size the lower the correlation BCS with reference series. For volatility measures (MCD 1, MCD 2) this pattern is not noticeable. Eventually this outcome cannot be treated as any prospective finding. The natural explanation seems to be the extremely low number of cases which represent such a relation. Optional case by case examination could bring better explanation but rather there is no chance for general and useful findings.

3. Summary and conclusions

Assuming sufficient differentiation among diverse sample sizes, in terms of quality of measurement some analyses were conducted.

Both sample size and effective sample size seem to have impact on quality measures especially as a trend. What is important, although tendencies are slight they are practically always coherent for all quality of measures. The key finding shows that these tendencies are observed especially in great part of the range of sample sizes and ESS. Along with further increase of sample size modest improvement of data quality is observed.

Having no arguments for strong statistical conclusions some general findings can be given.

- In general the effect on data volatility and consistency with reference series is clear as tendency (Chart 2.6 and 2.7).
- Concerning effective sample size the impact on MCD1, MCD2 is achieved also as a tendency, on part of the range quite stronger (Chart.2.12). As far as case of the impact on correlation BCS with reference series is concerned the tendency is hardly observed (Chart.2.11)

Gathering from GfK Polonia experience it can be stated that determining sample size is always important and challenging question. As the goal is to find the optimal solution different ways of determining sample size are considered. In everyday practice of GfK Polonia the basic parameters of estimations are taken into consideration. Often published tables are accounted and sometimes these parameter are calculated. This is what is done for most of ad hoc surveys. When there is a history of similar projects we revise how properly a sample size was calculated. Perhaps it is the best way to optimize sample size. In such a case we have maximum information about nonresponse level, variability of some important variables and finally sampling error. A very important aspect which is always considered is chance of subgroup analyses. It is very often the case. This very obvious that researches want to learn not only the estimation of global parameters but also differences among different groups. That's why some consideration about interesting subgroups is strongly recommended. As far as face-to-face method is concerned the key issue in our common experience is impossibility of using the simple random sample (SRS) approach. This limitation is mainly caused by unacceptable level of costs necessary for conducting data collection. This is one of the reasons why CATI and on-line methods are often more efficient. Having proper sampling frame there is no point to reject SRS design. All formulas can be directly applied. Otherwise computation for complex designs are quite time consuming.

Another common problem which should be faced with is a variability assumed before survey. The degree of variability in examined attributes is often unknown that's why there is a need to conduct a survey. Nevertheless there could always be some hypotheses about linkage of the measured attribute(s) with some other attributes of which distributions in population are known to some extent. Taking all of these into consideration one can gather a lot of information about potential heterogeneity of a population in the examined area. Knowing this there should be a common practice to assess a level of heterogeneity in a population, even if no standard approaches in this subject are developed. Quite often it is a matter of relevant expertise which helps to judge in which direction sample size should be driven: do we need a larger sample expecting heterogeneity or can we remain on basic level.

All conclusions drawn from this report rather support our intuition than bring striking findings. For more theoretical purposes there is a way to go deeper in the analytical process. In order to do that, some more information could be obtained. For each country there is a need to confront practice of DG ECFIN Consumer Survey with the theory. However in some cases it could be really challenging as a sample designs their apply are very complex. Perhaps it is a right line to compare not sample sizes themselves but rather their suitability. There should be always concern about quality of data as a whole and the sample size is one of its determinant. Being aware of complex conditions determining particular size of samples there should be a potential concern about the adequate sample size considering limitations at appropriate level.

4. References

1. Glossary of Statistical Terms OECD
2. Särndal, Carl-Erik; Swensson, Bengt; Wretman, Jan (2003). *Model assisted survey sampling*. Springer. pp. 9–12. ISBN 978-0-387-40620-6. Retrieved 2 January 2011
3. Israel, Glen D. (1992) *Determining Sample Size*, Agricultural Education and Communication Department, University of Florida, IFAS Extension, PEOD6 (reviewed 2013)
4. Turner, Anthony G. "Sampling frames and master samples". United Nations Secretariat. Retrieved 12/11/2012.
5. Data source: *Metadata_checked_by_partners - Consumers.xlsx*
6. Cochran, W. G. (1963). *Sampling Techniques*, 2nd Ed., New York: John Wiley and Sons, Inc.
7. Kish, Leslie. (1965). *Survey Sampling*. New York: John Wiley and Sons, Inc.
8. Sudman, Seymour. (1976). *Applied Sampling*. New York: Academic Press.
9. Yamane, Taro. (1967). *Statistics: An Introductory Analysis*, 2nd Edition, New York: Harper and Row.