EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics and Information Society
Unit F-3: Living Conditions and Social Protection Statistics

DOC LC-HBS/08/09/EN

WORKING PARTY

LIVING CONDITIONS

10-12 June 2009

EUROSTAT-LUXEMBOURG

Room: Ampère

MINIMUM SAMPLE SIZES FOR THE NEXT HBS

Determination of minimum sample sizes for the 2010 round of data collection of the Household Budget Surveys (HBS) in the EU

Introduction

The 2005 round of data collection of the Household Budget Surveys (HBS) in the European Union (EU) is coming to an end. Micro-data files and/or aggregated tables have been collected by Eurostat for all the EU-27 countries plus Norway and Croatia thus far. A series of indicators derived from the HBS 2005 data is now available on Eurostat's website[1] (NewCronos Database) for all these 29 countries.

One of the primary goals of the next HBS wave, which is due to take place in 2010, is to strengthen data comparability between the different countries. In particular, a prerequisite for greater comparability of the HBS data is to satisfy a minimum level of accuracy in each country.

In preparation for the 2010 round of data collection, this document presents and discusses different approaches to determine minimum sample sizes for each country that would ensure HBS estimates are acceptably accurate.

However, one has to keep in mind that any approach to determine minimum sample sizes has to be a balance between what are considered as conflicting concerns:

- Statistical concerns: that is, the need to ensure a minimum level of accuracy both at country, but also at EU level, since one is producing aggregates at this level as well.
- Cost concerns: any recommendation regarding minimum sample sizes has to take into account all the survey costs entailed to meet the requirements.
- Political concerns: the HBS is inherently a national instrument and so nobody is in a position to impose anything on anybody. All the improvements and efforts towards harmonization in this survey have been done throughout the years in the belief that any extra effort that is done by every country will benefit the others but will also pay dividends in bringing everybody in a position where they can compare with and benchmark against others. Besides this, one has to remember that such a rich source of information (which has probably the highest burden in all the National Surveys) is used by other users besides its primary user, which is the Consumer Price Index. One cannot underestimate the benefits that accrue by policy makers, researchers and managers, besides the media and the public in general, taking more informed and better decisions by making use of the information that is gathered through the survey. It is in this light that one should look at the different options this document is proposing.

[1] Population and social conditions/Living conditions and welfare/Consumption expenditure of private households

Sampling designs – HBS 2005

All the Household Budget Surveys for the reference year 2005 are sample surveys of private households. Most of the participating countries drew a sample of households in a way that the probability of a household being selected is known (technically known as a probability design). In this way, the results can be reliably projected from the sample to the household reference population with known levels of certainty/precision, .i.e. standard errors and confidence intervals for survey estimates can be constructed.

On the other hand, non-probability schemes (Quota selection) were implemented in Czech Republic and Germany. Although this type of sampling is generally quicker and cheaper, there is no assurance that the selection of households is not biased and is representative of the whole population. This error can be reduced if the enumerators are knowledgeable enough to choose alternative households with the same characteristics as the ones which are not available. Unfortunately, not enough information is available to guarantee that this has actually happened and/or to what degree.

When choosing the sample one has to decide whether one chooses addresses or persons at the ultimate stage. The former implies that all the private households currently residing at a selected address are eligible for inclusion in the HBS. If one targets the person one normally includes all the household the sampled person belongs to. Moreover, many of the samples were stratified according to geographical dimensions. This improves the representativity of the samples by ensuring a minimum adequate size by regions.

The HBS data were weighted. Sample weights are needed to correct for imperfections in the sample that might lead to bias and also to rectify other departures between the sample and the reference population. The design weights are calculated for each sampled household as the inverse of its probability of selection into the sample.

Another issue which one has to contend with in such surveys is total non-response because some households, which are initially chosen, do not take part in the survey. The reasons for this may include: absence, refusal, incapacity of the interviewee to participate, etc. Non-response generally increases bias in sample estimates, particularly if the non-respondents have specific characteristics. In order to reduce this, the household design weights are inflated by applying correction factors which actually represent the inverse of the household response probabilities. Response probabilities are estimated by fitting a response model to the data.

Finally, most of the HBS countries also “calibrated” the design weights corrected for non-response to external data sources: this weight adjustment brought the sample estimates into agreement with known population characteristics (e.g., population totals by age group, gender, NUTS2 region…) and then increased the sample accuracy.

Achieved samples sizes – HBS 2005

As mentioned previously, all the Household Budget Surveys are sample surveys of private households. Thus, any statistic which is derived from HBS data experience sampling errors, that is, the results obtained for any single sample would be likely to vary slightly from the true values for the population. The difference between the estimates derived from the sample and the true population values is referred to as the sampling error. The amount of variation can generally be reduced by increasing the size of the sample: the higher the sample size, the better the accuracy.
The achieved sample sizes for the 2005 HBS wave are shown in the table below. For comparison purposes, the achieved sample sizes for the 2005 wave of EU Statistics on Income and Living Conditions (EU-SILC) are also presented.

Table 1: Achieved sample sizes (in households) – HBS 2005/EU-SILC 2005

Country	HBS 2005[2]	EU-SILC 2005[3]
Austria (AT)	8400	5148
Belgium (BE)	3550	5166
Bulgaria (BG)	2870	-
Croatia (HR)	2727	-
Cyprus (CY)	2990	3746
Czech Republic (CZ)	2965	4351
Denmark (DK)	2449	5957
Estonia (EE)	3432	4208
Finland (FI)	4007	11229
France (FR)	10240	9775
Germany (DE)	52217	13111
Greece (GR)	6555	5568
Hungary (HU)	9058	6927
Ireland (IE)	6884	6085
Italy (IT)	24107	22032
Latvia (LV)	3774	3846
Lithuania (LT)	7586	4441
Luxembourg (LU)	3202	3622

[2] Source : Quality Report of the Household Budget Surveys – 2005

[3] Source: EU-SILC Comparative Quality Report for 2005

Country	HBS 2005[2]	EU-SILC 2005[3]
Malta (MT)	2586	3459
The Netherlands (NL)	1570	9562
Norway (NO)	3376	5996
Poland (PL)	34767	16395
Portugal (PT)	10403	4615
Romania (RO)	33066	-
Slovakia (SK)	4710	5414
Slovenia (SI)	3725	8287
Spain (ES)	8881	13027
Sweden (SE)	2079	6133
United Kingdom (UK)	6785	10826
EU-15	**151329**	**131856**
EU-25	**226922**	**192930**
EU-27	**262858**	**-**
NMS-10	**75593**	**61074**
Total	**268961**	**198926**

Although the Household Budget Surveys in the EU appear to have achieved higher sample sizes at EU-15 and EU-25 level than EU-SILC, which is actually the EU reference source for statistics on income and living conditions, there are important variations in the HBS sample sizes from one country to another: while the DE, IT, PL and RO samples comprise more than 20000 households and the samples for France and Portugal have more than 10000 households, the achieved sample sizes fall below 4000 in nearly half of the HBS countries (BE, BG, HR, CY, CZ, DK, EE, LV, LU, MT, NL, NO, SI and SE). Sample sizes of less than 4000 households in certain countries might be considered too low to draw reliable estimates from the HBS data for those countries.

Table 2: Achieved sample sizes – HBS 2005/EU-SILC 2005 (Summary statistics)

	Min	Max	Mean	CV (%)
HBS 2005	1570	52217	9275	128
EU-SILC 2005	3459	22032	7651	59

While the mean sample size for HBS 2005 (9275 households) is higher than for EU-SILC (7651), the coefficient of variation (CV), which measures variability in relation to the mean, is twice higher for HBS than for EU-SILC.

The above could seriously hamper EU comparability of the HBS data. For instance, one might question whether or not it is statistically meaningful comparing the data for Germany

(sample size = 52217) with the data for the Netherlands (sample size = 1570). Furthermore, if one had to consider the Design Effect (*Deff*), that is, the ratio of the variance under the actual sampling plan to the variance that would be obtained under simple random sampling of same size, the effective sample sizes, that is, the achieved sample sizes divided by *Deff*, are even smaller.

Table 3: Effective sample sizes (in households) – HBS 2005[4]

Country	Achieved Sample Size	*Deff*	Effective sample Size
AT	8400	2.28	3678
BE	3550	2.40	1477
BG	2870	1.00	2870
CY	2990	1.02	2935
CZ*	2965	1.66	1786
DE	52217	1.47	35453
DK	2449	3.14	780
EE	3432	2.45	1399
ES	8881	2.40	3702
FI	4007	1.66	2414
FR	10240	1.02	10061
GR	6555	1.37	4801
HR	2727	1.13	2422
HU	9058	2.01	4501
IE	6884	1.23	5601
IT*	24107	1.66	14522
LT	7586	2.02	3753
LU	3202	0.90	3563
LV	3774	2.17	1740
MT*	2586	1.66	1558
NL	1570	1.93	815
NO	3376	1.18	2867
PL*	34767	1.66	20944
PT*	10403	1.66	6267
RO	33066	1.93	17107
SE	2079	1.13	1842
SI	3725	1.54	2416
SK	4710	0.97	4852
UK	6785	1.40	4851

*For CZ, IT, MT, PL and PT, in the absence of a Deff estimate, the mean value of the Deffs for the other countries (Deff=1,66) was given

[4] Source : Quality Report of the Household Budget Surveys – 2005

To sum up, the problem posed by the HBS sample sizes is that of their distribution among the different countries. As mentioned, this has evident implications for comparability of the HBS data across the EU. A distribution as such is also likely to damage the accuracy at EU level. In fact, it is well-established optimality at EU level requires at least in each country sample sizes which are "in proportion to" the country size: the more populated the country is, the higher the sample size must be. Clearly, the achieved HBS sample sizes do not comply with this requirement: for instance, the sample sizes in FR, ES and UK, ones of the biggest EU countries, are relatively low (respectively 10240, 8881 and 6785 households) compared to the achieved sizes in less populated countries like, for instance, Austria (8400), Portugal (10403) or Romania (33066)

In order mainly to improve data comparability across the EU, alternative allocations of the HBS sample among the countries are most desirable for the next HBS wave. The so-defined sample sizes would constitute minimum sample sizes that should be achieved by the countries for the 2010 round of data collection of the HBS. As already mentioned in the introduction, opposite concerns have to be taken into account at this stage: on the one hand, the need to ensure a minimum level of accuracy both at EU and national level, on the other, the need for the minimum sample sizes not to be too burdensome for the countries. The latter means the achieved sample sizes for HBS 2005 should be taken into account somehow when determining the minimum sizes for HBS 2010.

Basically, two different approaches for determining minimum sample sizes are going to be explored. The first approach (called "*Optimality at country level*") sets out a minimum level of accuracy all the countries should achieve in 2010. The minimum sample size needed to meet the precision requirement is then derived for each country separately using analytic variance formulae. The second approach (called "*Optimality at EU level*") considers it as a pure problem of sample allocation among countries under the constraint of a fixed size at EU level. The latter problem is similar to a problem of sample allocation in case of stratified sampling.

Optimality at country level

Let $\overline{Y}$ be the mean total household consumption expenditure. Assuming the indicator follows a normal distribution, the relative confidence interval at 95% level is given by:

$$CI\left(\overline{Y}\right) \cong 1.96 \times \sqrt{Deff} \times \frac{CV}{\sqrt{n}} \qquad (1)$$

Where:

- *n* is the achieved household sample size.
- *CV* is the coefficient of variation over the household population of the variable "total consumption expenditure".
- *Deff* is the Design Effect, that is, the ratio of the variance of $\overline{Y}$ under the current sampling plan to the variance that would be obtained under simple random sampling of same size.
- *1.96* is the quantile at 97.5% of the normal distribution of mean 0 and standard deviation 1.

Hence, the minimum sample size needed to achieve a relative of precision of *x*% is:

$$n = Deff \times \left(1.96 \times \frac{CV}{x} \right)^2 \qquad (2)$$

In *(2)*, *Deff* and *CV* are unknown quantities because they refer to the household population that would be surveyed in the next HBS wave. Nevertheless, it is possible to work out estimates of these two quantities for each country using the HBS 2005 data. In the next table, minimum sample sizes using the HBS 2005 data are presented for different levels of relative precision: $x = 1, 2, 3, 4$ and 5%.

Table 4: Sample sizes for different levels of relative precision *x*

Country	Achieved sample size - HBS 2005	Deff[5]	Coefficient of Variation (%) of the total consumption expenditure[6]	x = 1%		x = 2%		x = 3%		x = 4%		x = 5%	
				Sample size	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*
AT	8400	2.28	68	40501	32101	10125	1725	4500	-3900	2531	-5869	1620	-6780
BE	3550	2.40	61	34307	30757	8577	5027	3812	262	2144	-1406	1372	-2178
BG	2870	1.00	54	11202	8332	2801	-69	1245	-1625	700	-2170	448	-2422
HR	2727	1.13	64	17781	15054	4445	1718	1976	-751	1111	-1616	711	-2016
CY	2990	1.02	70	19200	16210	4800	1810	2133	-857	1200	-1790	768	-2222
CZ*	2965	1.66	67	28627	25662	7157	4192	3181	216	1789	-1176	1145	-1820
DK	2449	3.14	56	37828	35379	9457	7008	4203	1754	2364	-85	1513	-936
EE	3432	2.45	85	68001	64569	17000	13568	7556	4124	4250	818	2720	-712
FI	4007	1.66	66	27778	23771	6945	2938	3086	-921	1736	-2271	1111	-2896
FR	10240	1.02	62	15062	4822	3766	-6474	1674	-8566	941	-9299	602	-9638
DE	52217	1.47	70	27671	-24546	6918	-45299	3075	-49142	1729	-50488	1107	-51110
GR	6555	1.37	69	25057	18502	6264	-291	2784	-3771	1566	-4989	1002	-5553
HU	9058	2.01	63	30647	21589	7662	-1396	3405	-5653	1915	-7143	1226	-7832
IE	6884	1.23	70	23153	16269	5788	-1096	2573	-4311	1447	-5437	926	-5958
IT*	24107	1.66	67	28627	4520	7157	-16950	3181	-20926	1789	-22318	1145	-22962
LV	3774	2.17	82	56053	52279	14013	10239	6228	2454	3503	-271	2242	-1532
LT	7586	2.02	71	39118	31532	9780	2194	4346	-3240	2445	-5141	1565	-6021
LU	3202	0.90	60	12447	9245	3112	-90	1383	-1819	778	-2424	498	-2704
MT*	2586	1.66	67	28627	26041	7157	4571	3181	595	1789	-797	1145	-1441

[5] Source: Quality Report of the Household Budget Surveys – 2005

[6] Source: HBS Summary Reports - 2005

NL	1570	1.93	52	20048	18478	5012	3442	2228	658	1253	-317	802	-768
NO	3376	1.18	66	19746	16370	4937	1561	2194	-1182	1234	-2142	790	-2586
PL*	34767	1.66	67	28627	-6140	7157	-27610	3181	-31586	1789	-32978	1145	-33622
PT*	10403	1.66	67	28627	18224	7157	-3246	3181	-7222	1789	-8614	1145	-9258
RO	33066	1.93	84	52315	19249	13079	-19987	5813	-27253	3270	-29796	2093	-30973
SK	4710	0.97	97	35061	30351	8765	4055	3896	-814	2191	-2519	1402	-3308
SI	3725	1.54	61	22014	18289	5503	1778	2446	-1279	1376	-2349	881	-2844
ES	8881	2.40	58	31016	22135	7754	-1127	3446	-5435	1938	-6943	1241	-7640
SE	2079	1.13	60	15628	13549	3907	1828	1736	-343	977	-1102	625	-1454
UK	6785	1.40	69	25606	18821	6401	-384	2845	-3940	1600	-5185	1024	-5761
EU-15	151329			393356	242027	98339	-52990	43706.21	-107623	24584.7	-126744	15734.2	-135595
EU-25	226922			749331	522409	187333	-39589	83258.95	-143663	46833.2	-180089	29973.2	-196949
EU-27	262858			812848	549990	203212	-59646	90316.43	-172542	50803	-212055	32513.9	-230344

*For CZ, IT, MT, PL and PT, the mean value of the CVs for the other countries (67%) and of the Deff (1,66) was given

The minimum sample sizes that are presented above for EU-15, EU-25 and EU-27 are just the summations of the minimum sample sizes over all the EU-15, EU-25 and EU-27 countries. However, it should be noted that a relative precision of *x*% in each country entails the same level of precision for estimates at EU level.

Looking at the numerical results, a relative precision of 3% for the mean total consumption expenditure seems to be a realistic objective for the next HBS round. Actually, except BE, CZ, DK, EE, LV, MT and NL, all the countries already reached the objective in 2005. As for the seven countries which have not reached it yet, they could make it for the next HBS wave by providing micro-data not only for one reference year, but for several years. Cumulating HBS data from several years will make the sample size higher and the accuracy better.

A variant approach to determine minimum sample sizes would be to use other benchmark indicators to assess the accuracy: for instance, instead of using the mean total consumption expenditure, one might use the mean consumption expenditure in a key COICOP category, for instance, CP04 (Housing, water, electricity, gas and other fuels) or CP07 (Transport)

Regardless of the benchmark indicator which is used, the approach implies that certain COICOP categories should be picked up as "strategic" groups for which a minimum level of accuracy is wanted. The choice of target groups is somewhat arbitrary and is always made at the expense of other groups which, in some particular cases, might happen to have a strategic interest too.

An alternative and probably more objective way to determine minimum sample sizes is to seek a relative precision of *x*% for all the COICOP categories that represent more than 100α% ($0<\alpha<1$) of the total expenditure. Using the same notations as in (*1*) and under some technical assumptions (see annex), the relative confidence interval at 95% level for the mean expenditure in a COICOP group which represent 100α% ($0<\alpha<1$) of the total expenditure is given by:

$$CI\left(\overline{Y}_{\alpha}\right) \cong 1.96 \times \sqrt{Deff} \times \frac{\sqrt{CV^2 + \left(1 + CV^2\right) \cdot \frac{1 + \left(\frac{1}{\alpha} - 1\right)^2}{2}}}{\sqrt{n}} \qquad (3)$$

Where:

- *n* is the achieved household sample size.
- *CV* is the coefficient of variation over the household population of the variable "total consumption expenditure".
- *Deff* is the Design Effect in relation to the mean total consumption expenditure.
- *1.96* is the quantile at 97.5% of the normal distribution of mean 0 and standard deviation 1.

The main lesson of *(3)* is the less aggregated the COICOP category is (i.e. the lower α is); the worse the level of relative precision is. Hence, the minimum sample size needed to achieve a relative of precision of *x*% is:

$$n = Deff \times \left(1.96 \times \frac{\sqrt{CV^2 + \left(1 + CV^2\right) \cdot \frac{1 + \left(\frac{1}{\alpha} - 1\right)^2}{2}}}{x} \right)^2 \qquad (4)$$

As in *(2)*, *Deff* and *CV* are unknown quantities which can be estimated using the data from HBS 2005. The next table presents minimum sample sizes as obtained under formula *(4)* for different levels of relative precision: x = 1, 2, 3, 4 and 5% and for different values of α: α = 0.5, 0.6, 0.7 and 0.8.

Table 5a: Sample sizes for different levels of precision *x* ($\alpha = 0.5$)

Country	Achieved sample size - HBS 2005	Deff	Coefficient of Variation (%) of the total consumption expenditure	x = 1%		x = 2%		x = 3%		x = 4%		x = 5%	
				Sample size	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*
AT	8400	2.28	68	81011	72611	20253	11853	9001	601	5063	-3337	3240	-5160
BE	3550	2.40	61	68623	65073	17156	13606	7625	4075	4289	739	2745	-805
BG	2870	1.00	54	22408	19538	5602	2732	2490	-380	1401	-1469	896	-1974
HR	2727	1.13	64	35566	32839	8891	6164	3952	1225	2223	-504	1423	-1304
CY	2990	1.02	70	38405	35415	9601	6611	4267	1277	2400	-590	1536	-1454
CZ	2965	1.66	67	57260	54295	14315	11350	6362	3397	3579	614	2290	-675
DK	2449	3.14	56	75669	73220	18917	16468	8408	5959	4729	2280	3027	578
EE	3432	2.45	85	136012	132580	34003	30571	15112	11680	8501	5069	5440	2008
FI	4007	1.66	66	55563	51556	13891	9884	6174	2167	3473	-534	2223	-1784
FR	10240	1.02	62	30129	19889	7532	-2708	3348	-6892	1883	-8357	1205	-9035
DE	52217	1.47	70	55348	3131	13837	-38380	6150	-46067	3459	-48758	2214	-50003
GR	6555	1.37	69	50119	43564	12530	5975	5569	-986	3132	-3423	2005	-4550
HU	9058	2.01	63	61302	52244	15325	6267	6811	-2247	3831	-5227	2452	-6606
IE	6884	1.23	70	46311	39427	11578	4694	5146	-1738	2894	-3990	1852	-5032
IT	24107	1.66	67	57260	33153	14315	-9792	6362	-17745	3579	-20528	2290	-21817
LV	3774	2.17	82	112115	108341	28029	24255	12457	8683	7007	3233	4485	711
LT	7586	2.02	71	78244	70658	19561	11975	8694	1108	4890	-2696	3130	-4456
LU	3202	0.90	60	24897	21695	6224	3022	2766	-436	1556	-1646	996	-2206
MT	2586	1.66	67	57260	54674	14315	11729	6362	3776	3579	993	2290	-296
NL	1570	1.93	52	40104	38534	10026	8456	4456	2886	2506	936	1604	34
NO	3376	1.18	66	39497	36121	9874	6498	4389	1013	2469	-907	1580	-1796
PL	34767	1.66	67	57260	22493	14315	-20452	6362	-28405	3579	-31188	2290	-32477
PT	10403	1.66	67	57260	46857	14315	3912	6362	-4041	3579	-6824	2290	-8113
RO	33066	1.93	84	104638	71572	26159	-6907	11626	-21440	6540	-26526	4186	-28880
SK	4710	0.97	97	70126	65416	17532	12822	7792	3082	4383	-327	2805	-1905
SI	3725	1.54	61	44033	40308	11008	7283	4893	1168	2752	-973	1761	-1964
ES	8881	2.40	58	62040	53159	15510	6629	6893	-1988	3878	-5003	2482	-6399
SE	2079	1.13	60	31260	29181	7815	5736	3473	1394	1954	-125	1250	-829
UK	6785	1.40	69	51217	44432	12804	6019	5691	-1094	3201	-3584	2049	-4736
EU-15	151329			786810	635481	196703	45374	87423	-63906	49176	-102153	31472	-119857
EU-25	226922			1498826	1271904	374706	147784	166536	-60386	93677	-133245	59953	-166969
EU-27	262858			1625872	1363014	406468	143610	180652	-82206	101617	-161241	65035	-197823

Table 5b: Sample sizes for different levels of precision *x* ($\alpha = 0.6$)

Country	Achieved sample size - HBS 2005	Deff	Coefficient of Variation (%) of the total consumption expenditure	x = 1%		x = 2%		x = 3%		x = 4%		x = 5%	
				Sample size	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*
AT	8400	2.28	68	69758	61358	17439	9039	7751	-649	4360	-4040	2790	-5610
BE	3550	2.40	61	59091	55541	14773	11223	6566	3016	3693	143	2364	-1186
BG	2870	1.00	54	19295	16425	4824	1954	2144	-726	1206	-1664	772	-2098
HR	2727	1.13	64	30626	27899	7656	4929	3403	676	1914	-813	1225	-1502
CY	2990	1.02	70	33070	30080	8268	5278	3674	684	2067	-923	1323	-1667
CZ	2965	1.66	67	49306	46341	12326	9361	5478	2513	3082	117	1972	-993
DK	2449	3.14	56	65158	62709	16289	13840	7240	4791	4072	1623	2606	157
EE	3432	2.45	85	117120	113688	29280	25848	13013	9581	7320	3888	4685	1253
FI	4007	1.66	66	47845	43838	11961	7954	5316	1309	2990	-1017	1914	-2093
FR	10240	1.02	62	25944	15704	6486	-3754	2883	-7357	1621	-8619	1038	-9202
DE	52217	1.47	70	47660	-4557	11915	-40302	5296	-46921	2979	-49238	1906	-50311
GR	6555	1.37	69	43158	36603	10789	4234	4795	-1760	2697	-3858	1726	-4829
HU	9058	2.01	63	52787	43729	13197	4139	5865	-3193	3299	-5759	2111	-6947
IE	6884	1.23	70	39879	32995	9970	3086	4431	-2453	2492	-4392	1595	-5289
IT	24107	1.66	67	49306	25199	12326	-11781	5478	-18629	3082	-21025	1972	-22135
LV	3774	2.17	82	96542	92768	24135	20361	10727	6953	6034	2260	3862	88
LT	7586	2.02	71	67376	59790	16844	9258	7486	-100	4211	-3375	2695	-4891
LU	3202	0.90	60	21439	18237	5360	2158	2382	-820	1340	-1862	858	-2344
MT	2586	1.66	67	49306	46720	12326	9740	5478	2892	3082	496	1972	-614
NL	1570	1.93	52	34533	32963	8633	7063	3837	2267	2158	588	1381	-189
NO	3376	1.18	66	34011	30635	8503	5127	3779	403	2126	-1250	1360	-2016
PL	34767	1.66	67	49306	14539	12326	-22441	5478	-29289	3082	-31685	1972	-32795
PT	10403	1.66	67	49306	38903	12326	1923	5478	-4925	3082	-7321	1972	-8431
RO	33066	1.93	84	90104	57038	22526	-10540	10012	-23054	5631	-27435	3604	-29462
SK	4710	0.97	97	60386	55676	15096	10386	6710	2000	3774	-936	2415	-2295
SI	3725	1.54	61	37917	34192	9479	5754	4213	488	2370	-1355	1517	-2208
ES	8881	2.40	58	53422	44541	13356	4475	5936	-2945	3339	-5542	2137	-6744
SE	2079	1.13	60	26917	24838	6729	4650	2991	912	1682	-397	1077	-1002
UK	6785	1.40	69	44103	37318	11026	4241	4900	-1885	2756	-4029	1764	-5021
EU-15	151329			677517	526188	169379	18050	75280	-76049	42345	-108984	27101	-124228
EU-25	226922			1290633	1063711	322658	95736	143404	-83518	80665	-146257	51625	-175297
EU-27	262858			1400032	1137174	350008	87150	155559	-107299	87502	-175356	56001	-206857

Table 5c: Sample sizes for different levels of precision *x* ($\alpha = 0.7$)

Country	Achieved sample size - HBS 2005	Deff	Coefficient of Variation (%) of the total consumption expenditure***	x = 1%		x = 2%		x = 3%		x = 4%		x = 5%	
				Sample size	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*
AT	8400	2.28	68	64476	56076	16119	7719	7164	-1236	4030	-4370	2579	-5821
BE	3550	2.40	61	54617	51067	13654	10104	6069	2519	3414	-136	2185	-1365
BG	2870	1.00	54	17834	14964	4459	1589	1982	-888	1115	-1755	713	-2157
HR	2727	1.13	64	28307	25580	7077	4350	3145	418	1769	-958	1132	-1595
CY	2990	1.02	70	30566	27576	7642	4652	3396	406	1910	-1080	1223	-1767
CZ	2965	1.66	67	45573	42608	11393	8428	5064	2099	2848	-117	1823	-1142
DK	2449	3.14	56	60224	57775	15056	12607	6692	4243	3764	1315	2409	-40
EE	3432	2.45	85	108252	104820	27063	23631	12028	8596	6766	3334	4330	898
FI	4007	1.66	66	44223	40216	11056	7049	4914	907	2764	-1243	1769	-2238
FR	10240	1.02	62	23979	13739	5995	-4245	2664	-7576	1499	-8741	959	-9281
DE	52217	1.47	70	44051	-8166	11013	-41204	4895	-47322	2753	-49464	1762	-50455
GR	6555	1.37	69	39890	33335	9972	3417	4432	-2123	2493	-4062	1596	-4959
HU	9058	2.01	63	48790	39732	12197	3139	5421	-3637	3049	-6009	1952	-7106
IE	6884	1.23	70	36859	29975	9215	2331	4095	-2789	2304	-4580	1474	-5410
IT	24107	1.66	67	45573	21466	11393	-12714	5064	-19043	2848	-21259	1823	-22284
LV	3774	2.17	82	89232	85458	22308	18534	9915	6141	5577	1803	3569	-205
LT	7586	2.02	71	62275	54689	15569	7983	6919	-667	3892	-3694	2491	-5095
LU	3202	0.90	60	19815	16613	4954	1752	2202	-1000	1238	-1964	793	-2409
MT	2586	1.66	67	45573	42987	11393	8807	5064	2478	2848	262	1823	-763
NL	1570	1.93	52	31918	30348	7979	6409	3546	1976	1995	425	1277	-293
NO	3376	1.18	66	31435	28059	7859	4483	3493	117	1965	-1411	1257	-2119
PL	34767	1.66	67	45573	10806	11393	-23374	5064	-29703	2848	-31919	1823	-32944
PT	10403	1.66	67	45573	35170	11393	990	5064	-5339	2848	-7555	1823	-8580
RO	33066	1.93	84	83282	50216	20820	-12246	9254	-23812	5205	-27861	3331	-29735
SK	4710	0.97	97	55814	51104	13953	9243	6202	1492	3488	-1222	2233	-2477
SI	3725	1.54	61	35046	31321	8761	5036	3894	169	2190	-1535	1402	-2323
ES	8881	2.40	58	49377	40496	12344	3463	5486	-3395	3086	-5795	1975	-6906
SE	2079	1.13	60	24879	22800	6220	4141	2764	685	1555	-524	995	-1084
UK	6785	1.40	69	40763	33978	10191	3406	4529	-2256	2548	-4237	1631	-5154
EU-15	151329			626217	474888	156554	5225	69580	-81749	39139	-112190	25049	-126280
EU-25	226922			1192909	965987	298227	71305	132545	-94377	74557	-152365	47716	-179206
EU-27	262858			1294025	1031167	323506	60648	143781	-119077	80877	-181981	51761	-211097

Table 5d: Sample sizes for different levels of precision *x* ($\alpha = 0.8$)

Country	Achieved sample size - HBS 2005	Deff	Coefficient of Variation (%) of the total consumption expenditure	x = 1%		x = 2%		x = 3%		x = 4%		x = 5%	
				Sample size	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*	*Sample size*	*Diff*
AT	8400	2.28	68	62022	53622	15505	7105	6891	-1509	3876	-4524	2481	-5919
BE	3550	2.40	61	52538	48988	13134	9584	5838	2288	3284	-266	2102	-1448
BG	2870	1.00	54	17155	14285	4289	1419	1906	-964	1072	-1798	686	-2184
HR	2727	1.13	64	27229	24502	6807	4080	3025	298	1702	-1025	1089	-1638
CY	2990	1.02	70	29403	26413	7351	4361	3267	277	1838	-1152	1176	-1814
CZ	2965	1.66	67	43838	40873	10959	7994	4871	1906	2740	-225	1754	-1211
DK	2449	3.14	56	57931	55482	14483	12034	6437	3988	3621	1172	2317	-132
EE	3432	2.45	85	104132	100700	26033	22601	11570	8138	6508	3076	4165	733
FI	4007	1.66	66	42539	38532	10635	6628	4727	720	2659	-1348	1702	-2305
FR	10240	1.02	62	23066	12826	5767	-4473	2563	-7677	1442	-8798	923	-9317
DE	52217	1.47	70	42374	-9843	10594	-41623	4708	-47509	2648	-49569	1695	-50522
GR	6555	1.37	69	38371	31816	9593	3038	4263	-2292	2398	-4157	1535	-5020
HU	9058	2.01	63	46932	37874	11733	2675	5215	-3843	2933	-6125	1877	-7181
IE	6884	1.23	70	35456	28572	8864	1980	3940	-2944	2216	-4668	1418	-5466
IT	24107	1.66	67	43838	19731	10959	-13148	4871	-19236	2740	-21367	1754	-22353
LV	3774	2.17	82	85836	82062	21459	17685	9537	5763	5365	1591	3433	-341
LT	7586	2.02	71	59904	52318	14976	7390	6656	-930	3744	-3842	2396	-5190
LU	3202	0.90	60	19061	15859	4765	1563	2118	-1084	1191	-2011	762	-2440
MT	2586	1.66	67	43838	41252	10959	8373	4871	2285	2740	154	1754	-832
NL	1570	1.93	52	30703	29133	7676	6106	3411	1841	1919	349	1228	-342
NO	3376	1.18	66	30239	26863	7560	4184	3360	-16	1890	-1486	1210	-2166
PL	34767	1.66	67	43838	9071	10959	-23808	4871	-29896	2740	-32027	1754	-33013
PT	10403	1.66	67	43838	33435	10959	556	4871	-5532	2740	-7663	1754	-8649
RO	33066	1.93	84	80112	47046	20028	-13038	8901	-24165	5007	-28059	3204	-29862
SK	4710	0.97	97	53690	48980	13422	8712	5966	1256	3356	-1354	2148	-2562
SI	3725	1.54	61	33712	29987	8428	4703	3746	21	2107	-1618	1348	-2377
ES	8881	2.40	58	47497	38616	11874	2993	5277	-3604	2969	-5912	1900	-6981
SE	2079	1.13	60	23932	21853	5983	3904	2659	580	1496	-583	957	-1122
UK	6785	1.40	69	39212	32427	9803	3018	4357	-2428	2451	-4334	1568	-5217
EU-15	151329			602379	451050	150595	-734	66931	-84398	37649	-113680	24095	-127234
EU-25	226922			1147500	920578	286875	59953	127500	-99422	71719	-155203	45900	-181022
EU-27	262858			1244767	981909	311192	48334	138307	-124551	77798	-185060	49791	-213067

As in Table 4, the minimum sample sizes that are presented in the above tables for EU-15, EU-25 and EU-27 are just the summations of the minimum sample sizes over all the EU-15, EU-25 and EU-27 countries. However, it should be noted that a relative precision of *x*% in each country entails the same level of precision for estimates at EU level.

Looking at the numerical results, a relative precision of 4% might be sought for all the COICOPs that account for more than 60% of the total expenditure. Actually, except BE, CZ, DK, EE, LV, MT and NL, all the countries already reached the objective in 2005. As for the seven countries which have not reached it yet, they could make it for the next HBS wave by providing micro-data not only for one reference year, but for several years. Cumulating HBS data from several years will make the sample size higher and the accuracy better.
Even though this approach relies on more objective criteria to select the benchmark indicators for accuracy, it leads to more conservative sample sizes, which might turn out to be an issue in practice.

Optimality at EU level

Contrary to the approach developed in the previous section, whereby minimum sample sizes are worked out country by country, the method which is presented here does not intend to achieve a minimum level of accuracy for the HBS data in each country independently, but only to increase the quality of the EU sample as a whole by proposing alternative allocations among the countries. In practice, the achieved sample size at EU level for HBS 2005 (268961 households in EU-27 plus Norway and Croatia) remain the target sample size for HBS 2010. The latter size can indeed be considered as an "acceptable" sample size at EU level (see table 1)

However, those 268961 households should be distributed differently in HBS 2010 than they were in HBS 2005. It means for instance that the most populated countries (e.g., FR, ES and UK) should receive the highest sample sizes. Nevertheless, the relative importance of big countries should be moderated in order to ensure a minimum level of accuracy in small countries.

Many sample allocation formulae are available in statistical literature, most of them relying on statistical criteria ("Statistical allocations"). However, in the absence of legal basis for HBS at EU level, cost criteria also has to be taken into account: this means balanced allocations should be sought for HBS 2010 by including the achieved sample sizes for HBS 2005.

Basically, three main statistical allocations can be considered:

1. Proportionate allocation

Proportionate allocation allocates a sample of n households using a sampling fraction in each of the countries that is proportional to that of the total household population. Let n_h denote the sample size in country h and N_h the country size (in number of households). Proportionate allocation implies:

$$n_h = n \cdot \frac{N_h}{\sum_h N_h} \qquad (5)$$

Relative to taking a completely unstratified sample, taking a proportionate sample is either a good thing, in that in reduced standard errors, or a neutral thing, in that standard errors don't change. Proportionate stratification can never increase standard errors. The reasoning is as follows:

- Total sampling variance can be decomposed into two components: within-strata variation and between-strata variation (the split between the two depending on how the strata are defined)
- Proportionate stratification decreases the between-strata variance. So, proportionate stratification is most efficient when the stratifiers that are used split the total variance in a way that maximizes the between-strata variance.

The above makes proportionate allocation quite interesting from a practical point of view, since it always makes the accuracy better, whatever the study variable. On the contrary, the Neyman allocation (see section below), though it achieves the smallest possible standard error of a given estimate, can happen to make the accuracy of other estimates worse.

2. Optimal (Neyman) allocation

The sample is allocated so as to achieve for a given size of sample the smallest possible standard error of estimate of the mean total consumption expenditure, i.e. the highest precision of the estimate. Distribution of the samples is weighted by the product of the country size and the within country standard deviation of the variable "total consumption expenditure". Let n_h denote the sample size in country h, N_h the country size (in number of households) and S_h the standard deviation of the variable "total consumption expenditure" over the whole household population. Neyman allocation implies:

$$n_h = n \cdot \frac{N_h S_h}{\sum_h N_h S_h} \qquad (6)$$

3. Compromise allocation

The Neyman allocation achieves the smallest possible standard error at EU level. Unfortunately, with this approach, the countries are treated in an unequal way: the smallest countries (N_h small) have the least precise results. In fact, the Neyman optimality is of an overall nature (here EU): it is the best strategy to produce EU results, but not country results.
Instead of minimizing an "EU" variance, the following criterion can be used:

$$\sum_{h=1}^{H} \left[(N_h)^{\alpha} CV\left(\hat{\bar{Y}}_h\right) \right]^2 \qquad (7)$$

Where:

- $\hat{\overline{Y}}_h$ is the mean total consumption expenditure in country *h*.
- N_h is the household population size of country *h*.
- α is a real and known fixed value, between 0 and 1.

$CV\left(\hat{\overline{Y}}_h\right)$ is a measure of imprecision within the country *h*, and the $(N_h)^\alpha$ is a weight which puts into perspective this measure. The overall EU quality criterion is obtained by weighting the country qualities by the importance of the countries. The importance is measured by $(N_h)^\alpha$. But, the exponent α comes to moderate the relative importance given to a country compared to the others.

Let n_h denote the sample size in country *h*, N_h the country size, S_h the standard deviation of the variable "total consumption expenditure" and $\overline{Y}_h$ the mean total expenditure in country *h*. (7) implies:

$$n_h = n \cdot \frac{(N_h)^\alpha \frac{S_h}{\overline{Y}_h}}{\sum_h (N_h)^\alpha \frac{S_h}{\overline{Y}_h}} \qquad (8)$$

The allocation formulae (5), (7) and (8) contain unknown quantities, namely the country size N_h, the standard deviation S_h and the mean expenditure $\overline{Y}_h$. However, all these quantities can be estimated using the HBS 2005 data.

Table 6: Sample sizes under different allocation schemes

Country	Achieved sample size - HBS 2005	Population size Nh (in households) - HBS 2005 (**)	CV (%) of the total consumption expenditure (***)	Standard deviation Sh (*1000)	Proportional allocation			Neyman allocation			Compromise allocation (α=0.5)		
					Sample size	Diff	Relative precision (%)	Sample size	Diff	Relative precision (%)	Sample size	Diff	Relative precision (%)
AT	8400	3490000	68	2069.1	4756	-3644	2.9	5980	-2420	2.6	8257	-143	2.2
BE	3550	1753696	61	1922.8	2390	-1160	3.8	2792	-758	3.5	5251	1701	2.6
BG	2870	3092051	54	163.6	4213	1343	1.6	419	-2451	5.2	6172	3302	1.3
HR	2727	1561250	64	736.0	2127	-600	2.9	952	-1775	4.3	5198	2471	1.8
CY	2990	250538	70	2159.9	341	-2649	7.5	448	-2542	6.5	2278	-712	2.9
CZ*	2965	4014512	67	478.8	5470	2505	2.3	1592	-1373	4.2	8726	5761	1.8
DK	2449	2553324	56	1861.5	3479	1030	3.3	3936	1487	3.1	5817	3368	2.6
EE	3432	557148	85	589.6	759	-2673	9.5	272	-3160	15.8	4124	692	4.1
FI	4007	2455000	66	1960.5	3345	-662	2.9	3986	-21	2.6	6722	2715	2.0
FR	10240	24918383	62	1837.2	33954	23714	0.7	37908	27668	0.6	20118	9878	0.9
DE	52217	37286612	70	2046.2	50807	-1410	0.7	63179	10962	0.7	27784	-24433	1.0
GR	6555	3992965	69	1868.6	5441	-1114	2.1	6178	-377	2.0	8962	2407	1.7
HU	9058	3837087	63	423.0	5228	-3830	2.4	1344	-7714	4.8	8022	-1036	2.0
IE	6884	1445414	70	3143.6	1970	-4914	3.4	3763	-3121	2.5	5470	-1414	2.1
IT*	24107	23571394	67	1879.6	32119	8012	0.9	36686	12579	0.9	21144	-2963	1.2
LV	3774	885029	82	490.4	1206	-2568	6.8	359	-3415	12.5	5014	1240	3.3
LT	7586	1461897	71	362.7	1992	-5594	4.4	439	-7147	9.4	5580	-2006	2.6
LU	3202	177910	60	3165.2	242	-2960	7.2	466	-2736	5.2	1645	-1557	2.8
MT*	2586	140093	67	1261.5	191	-2395	12.2	146	-2440	14.0	1630	-956	4.2
NL	1570	7090965	52	1578.7	9662	8092	1.4	9270	7700	1.5	9001	7431	1.5
NO	3376	2154179	66	2661.6	2935	-441	2.6	4748	1372	2.0	6297	2921	1.8

PL*	34767	13300839	67	430.7	18124	-16643	1.3	4743	-30024	2.5	15883	-18884	1.3
PT*	10403	3769096	67	1179.7	5136	-5267	2.4	3682	-6721	2.8	8455	-1948	1.8
RO	33066	7365336	84	240.5	10036	-23030	2.3	1467	-31599	6.0	14818	-18248	1.9
SK	4710	1900334	97	577.3	2589	-2121	3.7	909	-3801	6.2	8692	3982	2.0
SI	3725	697113	61	1082.0	950	-2775	4.8	625	-3100	5.9	3311	-414	2.6
ES	8881	14422209	58	1373.6	19652	10771	1.3	16404	7523	1.4	14318	5437	1.5
SE	2079	3883911	60	1793.1	5292	3213	1.7	5767	3688	1.6	7686	5607	1.4
UK	6785	25357064	69	2405.3	34552	27767	0.9	50504	43719	0.7	22585	15800	1.1
EU-15	151329				212797	61468	0.35	250499	99170	0.32	173216	21887	0.41
EU-25	226922				249649	22727	0.34	261376	34454	0.31	236476	9554	0.40
EU-27	262858				263898	1040	0.33	263262	404	0.31	257466	-5392	0.40
EU-27 + NO,HR	268961				268961	0	0.33	268961	0	0.31	268961	0	0.39

* For CZ, IT, MT, PL and PT, the mean value of the CVs for the other countries (67%) and of the Deff (1.66) was given

** Source: HBS Summary Reports (AT,BE,HR,DK,EE,FI,FR,DE,GR,HU,IE,LV,LT,RO,SK,ES and SE); EU-SILC 2005 (CY,CZ,IT,LU,MT,NL,NO,PL,PT,SI and UK) and Demographic Sources (BG)

*** Source: HBS Summary Reports - 2005

Another way to moderate the importance given to the biggest countries is to ensure a minimum sample size for all the countries. Let n be the total number of households (268961) to allocate among H countries. The approach consists of equally allocating a fixed proportion k ($0<k<1$) of these households. As for the $n(1-k)$ other households, they can be allocated among the countries using one the schemes we just presented (proportionate, optimal or compromise allocation). Finally, the sample size n_h which is received by country h is given by:

For proportionate allocation:

$$n_h = \overline{n} \cdot k + \overline{n} \cdot (1-k) \cdot \frac{N_h}{\sum_h N_h} = \overline{n} \cdot \left[k + (1-k) \cdot \frac{N_h}{\sum_h N_h} \right] \qquad (9)$$

For optimal allocation:

$$n_h = \overline{n} \cdot \left[k + (1-k) \cdot \frac{N_h S_h}{\sum_h N_h S_h} \right] \qquad (10)$$

For compromise allocation:

$$n_h = \overline{n} \cdot \left[k + (1-k) \cdot \frac{(N_h)^\alpha \frac{S_h}{\overline{Y}_h}}{\sum_h (N_h)^\alpha \frac{S_h}{\overline{Y}_h}} \right] \qquad (11)$$

T

The factor $\overline{n} = \frac{n}{H}$ is the mean household size for each country. The parameter k determines the relative importance given to country level estimation (which requires more equal sample sizes) versus EU level estimation. Actually, the formulae (9), (10) and (11) ensure a minimum sample size of $\overline{n} \cdot k$ households per country.

Table 7: Sample sizes under different allocation schemes and under the constraint of a minimum sample size for each country (*k*=0.25)

Country	Achieved sample size - HBS 2005	Population size Nh (in households) - HBS 2005	CV (%) of the total consumption expenditure	Standard deviation Sh (*1000)	Proportional allocation			Neyman allocation			Compromise allocation (α=0.5)		
					Sample size	Diff	Relative precision (%)	Sample size	Diff	Relative precision (%)	Sample size	Diff	Relative precision (%)
AT	8400	3490000	68	2069.1	5885	-2515	2.6	6803	-1597	2.4	8512	112	2.2
BE	3550	1753696	61	1922.8	4111	561	2.9	4413	863	2.8	6257	2707	2.3
BG	2870	3092051	54	163.6	5479	2609	1.4	2633	-237	2.1	6948	4078	1.3
HR	2727	1561250	64	736.0	3914	1187	2.1	3032	305	2.4	6217	3490	1.7
CY	2990	250538	70	2159.9	2575	-415	2.7	2655	-335	2.7	4027	1037	2.2
CZ	2965	4014512	67	478.8	6421	3456	2.1	3512	547	2.9	8863	5898	1.8
DK	2449	2553324	56	1861.5	4928	2479	2.8	5270	2821	2.7	6681	4232	2.4
EE	3432	557148	85	589.6	2888	-544	4.9	2523	-909	5.2	5412	1980	3.5
FI	4007	2455000	66	1960.5	4828	821	2.4	5308	1301	2.3	7360	3353	1.9
FR	10240	24918383	62	1837.2	27784	17544	0.7	30750	20510	0.7	17407	7167	0.9
DE	52217	37286612	70	2046.2	40424	-11793	0.8	49703	-2514	0.7	23157	-29060	1.1
GR	6555	3992965	69	1868.6	6399	-156	2.0	6952	397	1.9	9040	2485	1.7
HU	9058	3837087	63	423.0	6240	-2818	2.2	3327	-5731	3.0	8335	-723	1.9
IE	6884	1445414	70	3143.6	3796	-3088	2.5	5141	-1743	2.1	6421	-463	1.9
IT	24107	23571394	67	1879.6	26408	2301	1.0	29833	5726	1.0	18177	-5930	1.3
LV	3774	885029	82	490.4	3223	-551	4.2	2588	-1186	4.7	6079	2305	3.0
LT	7586	1461897	71	362.7	3813	-3773	3.2	2648	-4938	3.8	6504	-1082	2.5
LU	3202	177910	60	3165.2	2500	-702	2.2	2668	-534	2.2	3552	350	1.9
MT	2586	140093	67	1261.5	2462	-124	3.4	2428	-158	3.4	3541	955	2.8
NL	1570	7090965	52	1578.7	9565	7995	1.4	9271	7701	1.5	9069	7499	1.5

NO	3376	2154179	66	2661.6	4520	1144	2.1	5879	2503	1.8	7041	3665	1.7
PL	34767	13300839	67	430.7	15912	-18855	1.3	5876	-28891	2.2	14231	-20536	1.4
PT	10403	3769096	67	1179.7	6171	-4232	2.2	5080	-5323	2.4	8660	-1743	1.8
RO	33066	7365336	84	240.5	9846	-23220	2.3	3419	-29647	3.9	13432	-19634	2.0
SK	4710	1900334	97	577.3	4261	-449	2.9	3000	-1710	3.4	8838	4128	2.0
SI	3725	697113	61	1082.0	3031	-694	2.7	2787	-938	2.8	4802	1077	2.1
ES	8881	14422209	58	1373.6	17058	8177	1.3	14621	5740	1.5	13057	4176	1.5
SE	2079	3883911	60	1793.1	6288	4209	1.6	6644	4565	1.5	8083	6004	1.4
UK	6785	25357064	69	2405.3	28233	21448	1.0	40196	33411	0.8	19258	12473	1.2
EU-15	151329				194377	43048	0.37	222654	71325	0.34	164691	13362	0.45
EU-25	226922				245202	18280	0.36	253998	27076	0.33	235322	8400	0.43
EU-27	262858				260527	-2331	0.35	260049	-2809	0.33	255703	-7155	0.43
EU-27 + NO,HR	268961				268961	0	0.35	268961	0	0.32	268961	0	0.42

With *k*=0.25, the minimum sample size that is ensured for each country is:

$$n_{min} = \bar{n} \times 0.25 = \frac{268961}{29} \times 0.25 = 2318 \text{ households}$$

Table 8: Minimum sample size ensured for each country for different values of *k*

k	Minimum sample size ensured
0.1	927
0.2	1854
0.3	2782
0.4	3709
0.5	4637
0.6	5564
0.7	6492
0.8	7419
0.9	8347

In practice, the minimum sample size requirement n_{min} for each country is generally fixed in advance, and the value of *k* is determined accordingly:

$$k = \frac{n_{min}}{\bar{n}}$$

Statistical allocations, though they make the accuracy of the EU sample better, also imply huge costs for certain countries (e.g. FR, NL, ES, SE and UK) in order to reach the minimum sample size.

There is an evident need for a balance between statistical concerns and cost concerns. The latter means the difference between the achieved sample size for HBS 2005 and the target size that is wanted for HBS 2010 must be kept under control so only realistic efforts are required from the countries. A possible solution would be to use balanced allocations which include the achieved samples sizes for HBS 2005. The idea is to minimize the following criterion:

$$\sum_{h=1}^{H} \left[\left(\frac{N_h}{\sum_i N_i} \right)^{\alpha} \cdot \left(\frac{n_{0h}}{\sum_i n_{0i}} \right)^{\beta} \cdot CV\left(\hat{\bar{Y}}_h \right) \right]^2 \qquad (12)$$

Where:

- $\hat{\bar{Y}}_h$ is the mean total consumption expenditure in country *h*
- N_h is the household population size of country *h*
- n_{0h} is the achieved household sample size in country *h* for HBS 2005
- α and β are real and known fixed values, between 0 and 1.

$CV\left(\hat{\bar{Y}}_h\right)$ is a measure of imprecision within the country *h*, and the $(N_h)^{\alpha}$ is a weight which puts into perspective this measure. The term $(n_{0h})^{\beta}$ requires that the countries draw in 2010 "in proportion to" what they did for HBS 2005: thus, the minimum sample sizes for HBS 2010 are made not too burdensome for the countries because they take into account what the countries can effectively achieve. But, the exponent β means bigger efforts are required from the countries which achieved relatively low sample sizes in 2005
Using the same notations as previously, *(12)* is equivalent to:

$$n_h = n \cdot \frac{\left(\frac{N_h}{\sum_i N_i}\right)^{\alpha} \cdot \left(\frac{n_{0h}}{\sum_i n_{0i}}\right)^{\beta} \cdot \frac{S_h}{\bar{\bar{Y}}_h}}{\sum_h \left(\frac{N_h}{\sum_i N_i}\right)^{\alpha} \cdot \left(\frac{n_{0h}}{\sum_i n_{0i}}\right)^{\beta} \cdot \frac{S_h}{\bar{\bar{Y}}_h}} \qquad (13)$$

Table 9: Sample sizes under balanced allocation ($\alpha = \beta = 0.5$)

Country	Achieved sample size - HBS 2005	Population size Nh (in number of households)	Mean total consumption expenditure (EUR)	Standard deviation Sh (*1000)	BALANCED ALLOCATION			
					Sample size	Absolute Difference	Relative Difference (%)	Relative precision (%)
AT	8400	3490000	30428	2069.1	6922	-1478	-18	2.4
BE	3550	1753696	31521	1922.8	2861	-689	-19	3.5
BG	2870	3092051	3030	163.6	3024	154	5	1.9
HR	2727	1561250	11500	736.0	2483	-244	-9	2.7
CY	2990	250538	30856	2159.9	1139	-1851	-62	4.1
CZ	2965	4014512	7146	478.8	4346	1381	47	2.6
DK	2449	2553324	33241	1861.5	2633	184	8	3.8
EE	3432	557148	6936	589.6	2210	-1222	-36	5.5
FI	4007	2455000	29705	1960.5	3892	-115	-3	2.7
FR	10240	24918383	29632	1837.2	18619	8379	82	0.9
DE	52217	37286612	29232	2046.2	58069	5852	11	0.7
GR	6555	3992965	27081	1868.6	6637	82	1	1.9
HU	9058	3837087	6715	423.0	6983	-2075	-23	2.1
IE	6884	1445414	44909	3143.6	4151	-2733	-40	2.4
IT	24107	23571394	28053	1879.6	30027	5920	25	1.0
LV	3774	885029	5981	490.4	2817	-957	-25	4.5
LT	7586	1461897	5109	362.7	4445	-3141	-41	3.0
LU	3202	177910	52754	3165.2	851	-2351	-73	3.8
MT	2586	140093	18829	1261.5	758	-1828	-71	6.1
NL	1570	7090965	30360	1578.7	3262	1692	108	2.5
NO	3376	2154179	40328	2661.6	3346	-30	-1	2.4
PL	34767	13300839	6428	430.7	27087	-7680	-22	1.0
PT	10403	3769096	17607	1179.7	7887	-2516	-24	1.9
RO	33066	7365336	2863	240.5	24645	-8421	-25	1.5
SK	4710	1900334	5952	577.3	5456	746	16	2.5
SI	3725	697113	17738	1082.0	1848	-1877	-50	3.5
ES	8881	14422209	23682	1373.6	12341	3460	39	1.6
SE	2079	3883911	29885	1793.1	3205	1126	54	2.2
UK	6785	25357064	34859	2405.3	17015	10230	151	1.2
EU-15	151329				178373	27044	18	0.41
EU-25	226922				235463	8541	4	0.39
EU-27	262858				263132	274	0	0.39
EU-27 + NO,HR	268961				268961	0	0	0.39

Summary

Five main approaches to determine minimum sample sizes for HBS 2010 were presented in this document:

1. **Optimality at country level**

a. Use of the mean total expenditure as a benchmark for accuracy
b. Use of another COICOP group (for instance, Food and non-alcoholic beverages – CP01 –) as a benchmark for accuracy
c. Use as a benchmark of all the COICOP categories representing more than a fixed percentage of the total expenditure

2. **Optimality at EU level**

a. Use of a statistical allocation to define minimum sample sizes (Proportionate/Neyman/Compromise)
b. Use of an allocation which ensures a minimum sample size for each country
c. Use of a balanced allocation (inclusion of the achieved sample sizes for HBS 2005)

All these methods have pros and cons. Although the "*Optimality at country level*" methods ensure a minimum level of accuracy in each country, they also have inherent political implications in that, by treating each country separately, they "intrude" into HBS national matters. In the absence of legal basis for HBS at EU level, the latter might turn out to be quite difficult to handle.

With respect to this, the "*Optimality at country level*" methods appear to be more pragmatic in the sense that they present the problem as an EU problem. Nonetheless, the statistical gain from these methods should be less important than what we could hope using the former ones.

Annex

Let $\overline{Z}$ be the mean total consumption expenditure and let $\overline{Y}$ be the mean consumption expenditure in a given COICOP category *k*. It is assumed COICOP *k* accounts for 100α% of the total expenditure, that is, the ratio $\overline{Y}/\overline{Z}$ is equal to α.

1. The dispersion over the household population U of size N of the consumption expenditure in COICOP k is:

$$
\begin{aligned}
S_y^2 &= \frac{1}{N}\sum_{i\in U}\left(y_i - \overline{Y}\right)^2 \\
&= \frac{1}{N}\sum_{i\in U}\left(y_i - \alpha\cdot z_i + \alpha\cdot z_i - \overline{Y}\right)^2 \\
&= \frac{1}{N}\sum_{i\in U}\left(y_i - \alpha\cdot z_i\right)^2 + \frac{1}{N}\sum_{i\in U}\left(\alpha\cdot z_i - \overline{Y}\right)^2 + \frac{2}{N}\sum_{i\in U}\left(y_i - \alpha\cdot z_i\right)\left(\alpha\cdot z_i - \overline{Y}\right) \\
&= \frac{1}{N}\sum_{i\in U} z_i^2 \times\left(\frac{y_i}{z_i} - \alpha\right)^2 + \frac{1}{N}\sum_{i\in U}\left(\alpha\cdot z_i - \alpha\cdot \overline{Z}\right)^2 + \frac{2}{N}\sum_{i\in U}\left(y_i - \alpha\cdot z_i\right)\left(\alpha\cdot z_i - \alpha\cdot \overline{Z}\right) \\
&= \frac{1}{N}\sum_{i\in U} z_i^2 \times\left(\frac{y_i}{z_i} - \alpha\right)^2 + \alpha^2 S_z^2 + 2\alpha\cdot Cov\left(y_i - \alpha\cdot z_i, z_i\right)
\end{aligned}
$$

2. Assuming $Cov\left(y_i - \alpha\cdot z_i, z_i\right) \approx 0$, we have:

$$
\begin{aligned}
S_y^2 &\cong \frac{1}{N}\sum_{i\in U} z_i^2 \times\left(\frac{y_i}{z_i} - \alpha\right)^2 + \alpha^2 S_z^2 \\
&= \left(\frac{1}{N}\sum_{i\in U} z_i^2\right)\cdot\frac{\sum_{i\in U} z_i^2 \times\left(\frac{y_i}{z_i} - \alpha\right)^2}{\sum_{i\in U} z_i^2} + \alpha^2 S_z^2
\end{aligned}
$$

3. Since we have $0 \le \frac{y_i}{z_i} \le 1$ for all i, we obtain:

$$
\frac{\sum_{i\in U} z_i^2 \times\left(\frac{y_i}{z_i} - \alpha\right)^2}{\sum_{i\in U} z_i^2} \le \frac{\sum_{i\in U\left|\frac{y_i}{z_i}\ge\alpha\right.} z_i^2 \times(1-\alpha)^2 + \sum_{i\in U\left|\frac{y_i}{z_i}\le\alpha\right.} z_i^2 \times(\alpha)^2}{\sum_{i\in U} z_i^2} \cong \frac{\alpha^2 + (1-\alpha)^2}{2}
$$

4. Finally, the dispersion over the household population of the consumption expenditure in COICOP k is:

$$S_y^2 \cong \left(\frac{1}{N}\sum_{i\in U} z_i^2\right)\cdot \frac{\sum_{i\in U} z_i^2 \times \left(\frac{y_i}{z_i} - \alpha\right)^2}{\sum_{i\in U} z_i^2} + \alpha^2 S_z^2$$

$$= \left(S_z^2 + \bar{Z}^2\right)\frac{\alpha^2 + (1-\alpha)^2}{2} + \alpha^2 S_z^2$$

$$= S_z^2\left[\left(1 + \frac{1}{cv_z^2}\right)\cdot \frac{\alpha^2 + (1-\alpha)^2}{2} + \alpha^2\right]$$

5. The coefficient of variation over the household population of the consumption expenditure in COICOP *k* is given by:

$$cv_y^2 = \frac{S_y^2}{\bar{Y}^2} \cong \frac{S_z^2\left[\left(1 + \frac{1}{cv_z^2}\right)\cdot \frac{\alpha^2 + (1-\alpha)^2}{2} + \alpha^2\right]}{\alpha^2 \cdot \bar{Z}^2} = \left(1 + cv_z^2\right)\cdot \frac{1 + \left(\frac{1}{\alpha} - 1\right)^2}{2} + cv_z^2$$