Revision of the weighting strategy in the European Quality of Life Survey (EQLS)

Final report

Authors: José Vila and Jose Luis Cervera (DevStat)

Note: This report has not been subject to the standard Eurofound editorial procedure. It reflects the authors' opinions but not necessarily those of Eurofound.

Abstract

This report presents the main results of the independent revision of the weighting strategy of 3rd EQLS. The report assesses the design weights of 3rd EQLS and proposes a new calibration scenario that (1) applies the a Generalized Regression Estimation (GREG) method to compute the calibration weights and (2) considers a larger set of variables than that in the current calibration methodology of 3rd EQLS: annual activity status at country level (obtained from EU-SILC) and Education at NUTS 2 disaggregation level (obtained from LFS). A new treatment of the extreme weights is also proposed: instead of the current trimming strategy, extreme weights are managed through the addition of non-negativity constrains in the GREG algorithm and the application of a shrinkage method. The report also presents suggestions to be considered in 4th EQLS, such as the use of a non self-weighted sampling design.

Executive summary

Purpose

The quality assessment of 3rd EQLS found room for improvement in the calibration strategy applied to this survey. Specifically, the conclusions of the assessment report suggested that (1) design weights of EQLS should be revised and obtained as the inverse of the selection probabilities in each stratum; (2) additional calibration variables, specifically labour status, should be considered; and (3) a methodology to manage extreme weights that do not imply trimming rules should be developed.

The purpose of this revision of the weighting strategy is to follow these suggestions of the quality assessment report, developing and validating a new optimal weighting strategy, which could be applied in the next waves of EQLS.

Methodology

The methodology for the revision the weighting strategy of 3rd EQLS was threefold and included the assessment of the design weights, the proposal and validation of an optimal calibration scenario and the establishment of a new methodology to deal with extreme weights:

- The design weights were recovered from the description of the sampling design.

- Six alternative weighting scenarios were considered. Scenarios differed in both the additional recalibration variables (education level, activity status or both) and the calibration methodology (Generalized Regression Estimation methods – GREG- and Iterative Proportional Fitting - IPF). Alternative sources for the calibration variables were assessed. To select the optimal scenario, the six alternatives were compared according to four criteria: (1) level of distortion of the design weights, (2) impact on the sampling errors, (3) level of bias in those variables for which official data are available and are not included in the calibration procedure and (4) level of distortion of the estimations obtained with the current calibration methodology of EQLS.

- A combination of additional non-negativity constrains in the GREG model and a shrinkage approach were the proposed methodology for the management of extreme weights.

Findings

- ***Robustness with respect to the weighting strategy***. The application of the alternative weighting scenarios considered in the assessment does not generate relevant changes in the estimated distributions of EQLS variables. This robustness is a very positive feature of 3rd EQLS and, specifically, of its current sampling and calibration methodologies.

- ***The proposed calibration scenario*** consists of the application of GREG methodology more variables than those in the current calibration methodology of 3rd EQLS. Current calibration variables included number of households by household size at national level, as well as population by age cross with gender. The proposed calibration scenario also includes both the annual activity status at a country level (obtained from EU-SILC) and Education at NUTS 2 disaggregation level (obtained from LFS). EU-SILC is preferred to LFS as statistical source for the Activity Status due to the coherence in the definitions of this variable – based in self-assignment - in EU-SILC and EQLS.

- The addition of **non-negativity constrains** in the GREG algorithm and the application of a **shrinkage methodology** are useful procedures for the treatment of extreme weights. They allow for the elimination of potential negative weights arising in the optimal and reduce the extreme differences with the original design weight without introducing significant distortions in the estimates.

Conclusions and recommendations.

The analysis of the current weighting strategy of EQLS supports a series of recommendations that could be applied in the weighting of 4th EQLS:

- *Application of Generalized Regression Estimation methods* to compute the weights is recommended instead of Iterative Proportional Fitting. In EQLS, the latter methodology seems to generate a larger distortion of the design weights, without a relevant reduction of the bias.

- *Substitution of the current ad-hoc trimming strategy by a shrinkage method*, with a value of the shrinkage parameter is computed to minimize the mean square error of the estimates.

- *Self-weightiness need not be considered as a requirement in EQLS sampling.* 3rd EQLS sampling is designed in order to achieve self-weightiness. As a consequence of this requirement, too small samples need to be drawn in the smallest subpopulations. This fact may have an implication on both sampling errors and calibration procedures.

- *Specification of the design weights.* Independently of the final decision on considering or not self-weighted samples for 4th EQLS, it is necessary to record and disseminate all the information required for a detailed computation of the design weights.

- *Calibration by household size.* Calibration using household distribution by household size could be applied in future waves of EQLS, due to the lack of availability of updated statistics of population by household size. A higher degree in the harmonisation of the statistical sources to obtain the corresponding reference marginal distributions is recommended. EU-SILC provides with a reliable source for these distributions.

- *Dissemination of the design and calibrated population weights.* Eurofound's current strategy for dissemination of calibrated weights differs from other alternatives applied in other social surveys. For instance, ESocS disseminates population size weights, meanwhile Eurofound disseminates calibrated weights for the most standard combinations of countries. ESocS and EQLS practices could be integrated in 4th EQLS, with the dissemination of (1) weights for some basic aggregation of countries (as in 3rd EQLS) and (2) design, calibrated and population size weights (as in ESocS).

- *Harmonisation of the geo-codifications.* There are some other variables that could obviously play a role in calibration – such as the degree of urbanisation – and are related to geographical information (i. e. the municipality). To develop all the potential of such calibration variables, it is required a harmonisation of the geo-codifications used in EQLS with those in the ESS, specifically Eurostat's codifications applied in EU-SILC and LFS.

- *The role of Census data in EQLS calibration strategy* should consider the fundamental trade-off between accuracy versus timeliness and coherence that defines this type of data. The use of census data to calibrate some waves of EQLS and sampling-based data to calibrate other ones could have an impact on the time coherence of the data. For all these reasons, the use of census data to calibrate the 'dissemination' estimates of EQLS should only be considered under a staged approach. A first calibration stage using data from EU-SILC and LFS would provide with a series of estimations that are coherent along the different waves of EQLS. If census data were available in the future, Eurofound could update and disseminate the corresponding new weights. To this end, Eurofound should develop an in-house expertise to apply calibrations methods that can take advantage of the R programmes developed for this project.

- *Recalibration of 1st, 2nd and 3rd EQLS.* If any relevant modification in the weighting is considered for 4th EQLS and for the sake of time coherence, the dissemination of the results of the recalibration of the previous waves using the same methodology is highly recommended.

Table of Contents

Introduction 7

Chapter 1: Assessment of design weights 8

1.1. 3rd EQLS sampling design 8

1.2. Random routes for household selection 10

1.3. Is the sample of 3rd EQLS actually self-weighted? 10

1.4. Conclusions 17

Chapter 2: Development of alternative calibration scenarios 20

2.1. Calibration goals and potential calibration variables 20

2.2. Assessment of potential information sources for drawing calibration variables ... 22

2.3. Methods of calibration 26

2.4. Proposed calibration scenarios 27

Chapter 3: Implementation and comparison of the weighting scenarios 31

3.1. Implementation of the weighting scenarios 31

3.2. Criteria to compare weighting scenarios 31

Chapter 4: Strategies for treatment of extreme weights 52

4.1. Treatment of negative weights 52

4.2. Treatment of positive extreme weights 53

4.3. Conclusion 61

Chapter 5: Conclusions and recommendations 63

Introduction

DevStat Servicios de Consultoría Estadística S.L. (DevStat, hereinafter the Consultant) has been awarded the contract for the revision of the weighting strategy in the European Quality of Life Survey (EQLS).

This document is the final report of this project, which is structured as follows:

- **Chapter 1: Assessment of design weights**

  The first chapter assesses the design weights of 3rd EQLS and the hypothesis of the 3rd EQLS being actually self-weighted. The assessment is supported by a detailed analysis of the available information on the sampling procedure followed in each country.

- **Chapter 2: Development of alternative calibration scenarios**

  The second chapter evaluates potential variables, data sources and weighting methodologies that could be considered to define alternative calibration procedures. Finally, this section defines the six calibration scenarios to be analysed in Chapters 3 and 4.

- **Chapter 3: Implementation and comparison of the weighting scenarios**

  After computation of the corresponding weights for the six calibration scenarios, this chapter develops a comparison of the results according to four criteria: (1) level of distortion of the design weights, (2) impact on the sampling errors, (3) level of bias in those variables for which official data are available and are not included in the calibration procedure and (4) level of distortion of the estimations obtained with the current calibration methodology of EQLS.

- **Chapter 4: Strategies for treatment of extreme weights**

  Chapter 4 analyses the different impact of different methodologies to deal with negative and extreme weights and proposes a ‘shrinking’ procedure that could be applied in EQLS.

- **Chapter 5: Conclusions and recommendations**

  This final chapter presents a brief summary with the three main conclusions from the weighting exercise and suggests recommendations that could be applied for the weighting strategy of 4th EQLS.

Chapter 1: Assessment of design weights

The first step of calculating sample survey final weights is to calculate design weights as the inverse of each individual probability of selection. Then, when necessary, these weights are corrected using reweighting in order to, for example, solve missing value problems or fix population frame with information from external sources.

With regard to the treatment of non-response by reweighting procedures, basically two methods are commonly found. The first one is based on the assumption that non-response in each sample unit is random, therefore this method weights the effective sample in each sample unit to the whole theoretical sample of the unit by the inverse of its response ratio. The second method assumes that non-response depends on socio-demographic characteristics; therefore the results of the sample are weighted separately for each homogeneous group of final units. This approach is, at large extent, equivalent to reweighting the sample using as auxiliaries variables those variables that define such homogeneous. In this work, this is the proposed mechanism; so, non-response treatment and calibration will be done in a single step.

In the EQLS, design weights only consider the size of the household, assuming that the sample of households and/or individuals (depending on the country) is self-weighted i.e. the probability of selection is the same for all sampling units of the sample.

As a first step, a detailed analysis of each EQLS related countries sample plans has been carried out in order to verify whether the hypothesis of self-weighting is adjusted within the effective sampling design of the different countries involved.

1.1. 3rd EQLS sampling design

With some variations, the more general sampling design in EQLS consists of a three-stage sampling. A first stage at which sampling units (SU) are clusters of households (communities, electoral districts, etc.) that have been previously stratified. A second stage at which households are selected within the selected SU, and a third stage at which an adult is selected from each selected household.

According to this general scheme, design weights are obtained as the inverse of the selection probabilities at each stage. The probability of selecting a person of household j in SU i of stratum h, would be:

$$P(person_{hijk}) = P(SU_{hi}) * P(household_{hij}|SU_{hi}) * P(person_{hijk}|household_{hij}) \quad (1)$$

In the current weighting scheme of 3rd EQLS, in the weighting report, the only design weight that is considered is that derived from household size, that is, the last factor of the expression (1). This assumes, implicitly, that the result of the multiplication of the first two factors is constant for all units in the sample. Therefore, the probability of selection of households would be identical in each country and thus, the sample of households would be self-weighting.

A simple mechanism to achieve self-weighted samples of households is (1) at the first stage, to assign to each stratum a sample of primary sample units (PSU) proportional to the number of households (2) to select the sample of PSU in each stratum with probability proportional to the number of households, and (3) at the second stage, to select by simple random sampling a fixed number of households, say m, in each PSU.

For example, let N_h the number of households in stratum h, N the number of total households in the population and n the sample of PSU; N_{hi} the number of households in PSU_{hi} and k the fixed sample of households per PSU. Then the selection probability of the household j of PSU i in stratum h will be:

$$P(u_{hij}) = n * \frac{N_h}{N} * \frac{N_{hi}}{N_h} * \frac{k}{N_{hi}} = \frac{n * k}{N}$$

Therefore, the probability of selection is (n*k)/N for all households in the sample and the design weight will be a constant multiple of the number of adults in each household. This has been the most widely used method to generate self-weighted samples of households in the EQLS. The rest of the countries with self-weighted samples (Malta, The Netherlands and Sweden) are SRS of persons or households.

An alternative mechanism that also generates self-weighted samples of households is to select, in all strata, and at every stage a fixed proportion of units by simple random sampling. In the case of two stages, the probability of selection of a household will be f1*f2, with f1 being the proportion of PSU selected at the first stage and f2 the proportion of households selected at the second stage. If f1 and f2 are constant across strata, the probability will be constant too. For example, with SRS in the two stages, the case of a final and constant sampling proportion of n/N in every stratum is a particular case of constant f1 and f2 in each stratum, provided that f1*f2=n/N. Alternatively, if the sample fraction of the first stage varies between strata (f_{h1}, where h is the stratum), the sample fraction in the second stage (f_{h2}) should offset the variation, so that, in the end, $f_{h1}*f_{h2}$ = n/N across strata.

On the other hand, the types of sampling methodologies applied in EQLS that do not lead to self-weighting samples are much diverse and to establish a comprehensive typology is not an easy task. The most common reason for a deviation from having a self-weighted sample is a selection of PSU by SRS at first stage, followed by the selection a fixed number of households in each PSU at the second stage (as in Germany, Spain or Hungary). In this case, assuming that in stratum h there are M_h PSU and that m_h of them are chosen, and that in PSU_{hi} there are N_{hi} households, of which r are selected, the probability of selection of household u_{hij} would be:

$$P(u_{hij}) = P(PSU_{hi}) * P(u_{hij}|PSU_{hi}) = \frac{m_h}{M_h} * \frac{r}{N_{hi}}$$

That, although the proportion of PSU in each stratum is constant (i.e. $m_h/M_h$is constant), the probability depends on the selected PSU. If the size of PSU is roughly equal, the ratios r/M_h can be also considered roughly equal, but the factor m_h/N_{hi} still depends on the selected PSU_{hi}, so the product could be non-constant.

1.2. Random routes for household selection

Although, in principle, selection techniques based on random routes can be considered as probabilistic methods, in practice, if a strict control over the mechanics of selection is not maintained, they can develop complications and sources of bias (Kish, L.: Survey Sampling. John Wiley & Sons, Inc. New York, London 1965), that make it very difficult, if not impossible, to calculate the probability of selection of every possible sampling unit. When it is used, two common assimilations to usual probabilistic methods are commonly considered:

1. If the area (sample unit) where a given random route is developed is small enough to safely suppose that every unit in the area has the same approximate probability of being chosen, the selection could be though as equivalent to simple random sampling.
2. If some of the above requirements are not met, the random route should be considered as an additional stage, and so the units of a given random route will form a cluster.

In this report, the assessment if the EQLS samples are self-weighted will be carried out on the standard assumption that random routes, as a method of selection of households in each PSU, are equivalent to simple random sampling.

1.3. Is the sample of 3rd EQLS actually self-weighted?

Table 1.1. shows schematically the design of the sample in each of the countries. The sampling steps are described for each of them, by defining the sample units, the selection method and the allocation of the sample. At the first stage the stratification, if any, is further specified. The row headed self-weighted indicates whether, according to the design, the design generates a self-weighted sample of households.

It is important to note that, with some exceptions, the reports do not provide enough information to determine the sample design completely. Therefore, the content of the table should be considered as a proxy of the actual sampling methodology.[1]

[1][1]List of acronyms in table 1.1.

Selection	SRS	Simple Random Sampling (every unit has the same probability of being chosen)
	PPS	Each unit has Probability of being chosen Proportional to its Size
	RR	Households selected by means of Random Routes
	NB	Respondent in the household is selected by Next Birthday method
	SYS	Units are chosen by Systematic method with random start
Allocation	PPS	The number of selected units in each stratum is Proportional to the Size of the stratum
Units	HH	Households

Box 1.1. Remarks about table 1.1

Table 1.1. summarises a large amount of information to compare the sampling methodologies applied in all the countries. This box describes the meaning of the symbols and terms used in the table to facilitate its interpretation.

Question marks:

In these cases where some specific feature is not made clear by the explanations in the EQLS sampling report, a question mark is written beside consultant's interpretation of the information in the corresponding sampling report. For instance:

- In some countries, the selection method - as it is described in the sampling report - could be interpreted as PPS selection, although the corresponding explanation is not completely clear.

For example, in the case of Ireland it can be read "Electoral divisions will then be randomly selected in each stratum proportional to population". That could mean that selection is PPS, but also that sampling allocation in strata is proportional. As the second alternative is stated in point "2. Sample design" of the sampling report as common to all of the countries, maybe the additional information in the report of Ireland is most likely about PPS. In cases like that, the table shows "PPS?" and if PPS was considered true, the sample would be self-weighted.

- In other countries, the available information do not allow for determining whether the sample is self-weighted or not. In those cases the self-weighted cell is filled with a question mark.

Systematic sampling

The interpretation of systematic sampling depends on the mechanism of selection:

- When the units are selected out of a list which can be considered sorted at random, sampling has been assimilated to simple random sampling (i.e. Czech Republic).
- When each unit has a probability of selection proportional to its size. So, the result has been assimilated to sampling with probability proportional to size (i. e. Belgium or Greece).

Number of stages

In some cases, the sampling reports specify intermediate selection stages. This feature is presented in the row '#stages' with the symbol '+1'.

Random routes

According to paragraph 1.2, a selection trough random route has been assimilated to simple random sampling or as an additional stage. The latter case has been highlighted adding the "Type 2" to the abbreviation for random route (RR).

Table 1.1: Key features of 3rd EQLS sampling methodology by country

COUNTRY		AUSTRIA	BELGIUM	BULGARIA	CYPRUS	CZECH REPUBLIC	GERMAN Y	DENMARK	ESTONIA
RANDOM ROUTES		NO	NO	YES	YES	NO	YES	NO	YES
# STAGES		3+1	3+1	3	3	3	3	3	2
STAGE 1	STRATA	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	DIST X URB	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	COUNTY X SETTELMENT TYPE
	SAMPLING UNIT	MUNICIPALITIE S	MUNICIPALITIES	SETTELMENTS	ENUMERARION AREA	POSTAL DELIVERY DISTRICT	ADM AREA	500M RADIUS CIRCLES	HH
	SELECTION	SRS	PPS	PPS	PPS	SYS	SRS	PPS	RR
	ALLOCATION	PPS	PPS	PPS	PPS	PPS	PPS	PPS	PPS
STAGE 2	SAMPLING UNIT	POST CERTIFIED ADDRESS CODE	POST CODE CLUSTER	HH	HH	HH	HH	HH	PERSON
	SELECTION	SRS	SRS	RR	RR	SYS	RR	SRS	NB
	ALLOCATION	1	PPS	8	10	16	7	11	1
STAGE 3	SAMPLING UNIT	HH	HH	PERSON	PERSON	PERSON	PERSON	PERSON	
	SELECTION	SRS	SRS	NB	NB	NB	NB	NB	
	ALLOCATION	12	10	1	1	1	1	1	
		AUSTRIA	BELGIUM	BULGARIA	CYPRUS	CZECH REPUBLIC	GERMAN Y	DENMARK	ESTONIA
STAGE 4	SAMPLING UNIT	PERSON	PERSON						
	SELECTION	NB	NB						
	ALLOCATION	1	1						
SELF-WEIGHTED		NO	NO	YES	YES	NO	NO	NO	NO
NOTES								PSU are not disjoint sets	RR Type 2

COUNTRY		GREECE	SPAIN	FINLAND	FRANCE	HUNGARY	IRELAND	ITALY	LITHUANIA
RANDOM ROUTES		YES	YES	NO	YES	NO	NO	YES	YES
# STAGES		3	3	3+1	3	2	3	3	2+1
STAGE 1	STRATA	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	UDA Regions X URB	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	COUNTY X URBANISATION
	SAMPLING UNIT	Depends on area type	CENSUS AREAS	MUNICIPALITIES	COMMUNES	SETTELMENTS / DEPARTAMENTS	WARDS / ELECTORAL DIVISIONS	POSTAL CODES	RR
	SELECTION	PPS	SRS	PPS	PPS	SRS	PPS?	PPS?	SRS
	ALLOCATION	PPS	PPS	PPS	PPS	PPS	PPS	PPS	PPS
STAGE 2	SAMPLING UNIT	HH	HH	POSTCODE CLUSTER		PERSON	HH	HH	HH
	SELECTION	RR	RR	SYS	RR	SRS	SRS	RR	RR
	ALLOCATION	9	5	PPS	8	9	7	9	10
STAGE 3	SAMPLING UNIT	PERSON	PERSON	HH			PERSON	PERSON	PERSON
	SELECTION	NB	NB	SRS	SRS(HH)		NB	NB	NB
	ALLOCATION	1	1	20	3		1	1	1
STAGE 4	SAMPLING UNIT			PERSON					
		GREECE	SPAIN	FINLAND	FRANCE	HUNGARY	IRELAND	ITALY	LITHUANIA
	SELECTION			NB					
	ALLOCATION			1					
SELF-WEIGHTED		YES	NO	YES	NO	NO	YES	YES	NO

COUNTRY		LUXEMBOURG	LATVIA	MALTA	NETHERLAND	POLAND	PORTUGAL	ROMANIA	SWEDEN
RANDOM ROUTES		NO	NO	NO	NO	NO	YES	YES	NO
# STAGES		3	3	1	2	3	3	3	1
STAGE 1	STRATA	ELECTORAL DISTRICTS X URBANISATION	NUTS 2 X URB	NO	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB	NUTS 2 X URB
	SAMPLING UNIT	AD-HOC	CITIES /PARISHES	PERSON	POSTAL DELIVERY POINTS	GMINA	LOCALITIES	LOCALITIES	PERSON
	SELECTION	SRS	SRS	SRS	SRS	SRS	PPS	SRS	SRS
	ALLOCATION	PPS	PPS		PPS	PPS	PPS?	PPS	PPS
STAGE 2	SAMPLING UNIT	HH	HH		PERSON	HH	HH	HH	
	SELECTION	SRS	SRS		NB	SRS	RR	RR	
	ALLOCATION	5	10		1	6	7?	7?	
STAGE 3	SAMPLING UNIT	PERSON	PERSON			PERSON	PERSON	PERSON	
	SELECTION	NB	NB			NB	NB	NB	
	ALLOCATION	1	1			1	1	1	
STAGE 4	SAMPLING UNIT								
		LUXEMBOURG	LATVIA	MALTA	NETHERLAND	POLAND	PORTUGAL	ROMANIA	SWEDEN
	SELECTION								
	ALLOCATION								
SELF-WEIGHTED		NO	NO	YES	YES	NO	YES	NO	YES

COUNTRY		SLOVENIA	SLOVAKIA	UNITED KINGDOM	TURKEY	CROATIA	FYROM	KOSOVO
RANDOM ROUTES		NO	YES	NO	YES	YES	YES	YES
# STAGES		2	3	3	3	3	3	3
STAGE 1	STRATA	NUTS 2 X URB	DIST X URB	NUTS1 X URB	NUTS2 X URB	REGIONS X URB	NUTS3 X URB	DISTRICTS X URB
	SAMPLING UNIT	DISTRICTS	MUNICIPALITY	CENSUS SUPER OUTPUT AREAS	DISTRICTS	COUNTIES	ELECTORAL UNIT	ELECTORAL WARDS
	SELECTION	?	SRS	PPS	SRS	SRS	SRS	SRS
	ALLOCATION	PPS	PPS	PPS	PPS	PPS	PPS	PPS
STAGE 2	SAMPLING UNIT	HH	HH	HH	HH	HH	HH	HH
	SELECTION	SRS	RR	SRS	RR	RR	RR	RR
	ALLOCATION	10	10	9	16	10	10	10
STAGE 3	SAMPLING UNIT		PERSON	PERSON	PERSON	PERSON	PERSON	PERSON
	SELECTION		NB	NB	NB	NB	NB	NB
	ALLOCATION		1	1	1	1	1	1
STAGE 4	SAMPLING UNIT							
		SLOVENIA	SLOVAKIA	UNITED KINGDOM	TURKEY	CROATIA	FYROM	KOSOVO
	SELECTION							
	ALLOCATION							
SELF-WEIGHTED		NO	NO	YES	NO	NO	NO	NO

COUNTRY		SERBIA	MONTENEGRO	ICELAND
RANDOM ROUTES		YES	YES	NO
# STAGES		3	3	2
STAGE 1	STRATA	NUTS2 X URB	NUTS2 X URB	NUTS2 X URB
	SAMPLING UNIT	MUNICIPALITY	MUNICIPALITY	POSTCODE
	SELECTION	SRS	SRS	SRS
	ALLOCATION	PPS	PPS	PPS
STAGE 2	SAMPLING UNIT	HH	HH	PERSON
	SELECTION	RR	RR	SRS
	ALLOCATION	6	20	?
STAGE 3	SAMPLING UNIT	PERSON	PERSON	
	SELECTION	NB	NB	
	ALLOCATION	1	1	
STAGE 4	SAMPLING UNIT			
	SELECTION			
	ALLOCATION			
SELF-WEIGHTED		NO	NO	?

1.4. Conclusions

In general, the procedure of statistical calibration of sample surveys is to obtain a new set of sampling weights, from design weights, so that the estimators based on these new weights reproduce certain population totals that are known without or reduced error (CE Särndal. (2007), The Calibration Approach in Survey Theory and Practice. Survey Methodology, Vol 33, No. 2, 99-119). In addition, to maintain consistency with the sampling design, the new weights must be "as similar as possible" to the original design weights. This similarity is obtained, in practice, defining a distance measure between the two sets of weights.

The problem, thus presented, corresponds to one of minimization (the distance between design and calibrated weights) subject to some restrictions (the estimate of certain totals).

In theory of sample calibration, therefore, the role of design weights is crucial because they are the guarantee that the estimates obtained with the new weights are consistent with the original sample design and "inherit" the most important properties of original estimators (Deville, JC, and Särndal, CE (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association , 87, 376-382)

From the above discussion it is apparent that in most cases, samples of households are not self-weighted, or at least from the sampling reports it does not follow that they are. However the calibration procedure has been implemented using only the weights of household size, which would only be methodologically permissible in the case where the samples of households were self-weighted. Moreover, it is possible to calculate the original design weights in only a few cases, since the minimum information required for this computation is not in general available (see box 1.2).

Box 1.2 Information that should be included in any sampling report

The information available in the methodological reports and used by the consultant does not always allow for a complete understanding of how the sampling procedure has been implemented in all countries. In future waves of EQLS a standard sampling report should be required at country level, specifying at least:

- **Description of sampling stages**

In sampling surveys, balance of costs and effectiveness could lead to select units in two or more stages, which are each successive phases that are carried out in order to select the final sampled units. At each stage, the units that are to be chosen must be clearly defined, describing the set where they will be selected and its exact or approximate number and size. Additionally, it is important to specify the relation between every pair of successive stages.

- **Method of selection at each stage**

At each stage the method of selection of the units must be clearly described. This includes at least four types of information:

1) The stratification, if any, of units before selecting them, that is, the division of units in exhaustive and exclusive sets (therefore every unit of this stage should belong to one and only one of them). Besides that, the selection of units in one stratum must be independent of the selection in any other stratum.

2) The allocation of the sample in each stratum, that is, the number or proportion of units that will be selected in each stratum.

3) The effective procedure carried out to select each unit of the sample at each stage. The method must state, for every stratum, if the probability of selection is the same for all the units or if each unit has its own probability (for example, proportionate to its size) and if the units can be selected more than once, that is to say, if the selection is with or without replacement.

4) In order to take account the sampling complexities that random routes generate, if the sample includes this method of selection, is very important the sampling report describes perfectly the conditions under they are carried out.

- **Information to compute design weights**

Given that, in general, the sample design weights are derived from selection probabilities and sample proportions that, where appropriate, are derived from the stratification, the minimum information required to calculate the sample design weights should allow the calculation of both components. That is, in general, for each of the stages it should be known

1) Sample and population sizes of each strata

2) Selection probabilities of each SU of the sample

This information also must be possible to be linked to the microdata file in order to assign to each sample record the correct design weight. That is to say, each record in the sample should include stratum identifiers and the SU to which the record belongs at each stage[2].

In those cases where the appropriate registers were not available or they were not reliable enough to be used as a sampling frame, a random route approach could be applied. The specific sampling methodology for random routes needs to be customised to the specific characteristics of each country after discussion with the local fieldwork representatives of the contractor.

Example 1

1 Stage: Stratification and selection SRS of PSU in each stratum

2 Stage: Selection SRS of households in each selected PSU

3 Stage: Selection SRS of 1 adult in each selected household

In this case it would be necessary to know 1) the number of PSUs in the population and in the sample from each stratum 2) the number of households in the sample and in the population of each selected PSU and 3) the number of adults in selected households.

A variation of this case, particularly simple, is when at both stages the sampling rate is the same for all units. In this case is enough to know, besides the population sizes of the strata and the number of adults in the selected households, the sampling rate in each of the first two stages. If, in addition, samples are proportional to population of strata, it is not even necessary to know the population sizes of the strata but the total population.

[2] This linkage should have in consideration the standard requirements to guarantee anonymity. In these cases where the population size of a stratum is small enough to allow for an identification of the respondent, the information should be provided for a combination of strata.

Example 2

1 Stage: Stratification and selection of PSU in each stratum with probability proportional to the number of households in each PSU

2 Stage: Selection SRS of m households in each selected PSU

3 Stage: Selection SRS of 1 adult in each selected household

According to the discussion of this sampling mechanism that has been presented in section 1.1, in this case, since the number of households in each selected PSU is cancelled when calculating the probability of selection, it is only necessary to know 1) the number of households in the population and in the sample from each stratum 2) the fixed number m of households that are selected in each PSU and 3) the number of adults in the selected households.

Chapter 2: Development of alternative calibration scenarios

2.1. Calibration goals and potential calibration variables

Calibration has, in general, four goals[3]:

- Reduce the variance estimates correlated with total population marginal
- Improve longitudinal analysis, smoothing the oscillations introduced by the sample composition
- Reduce non-response bias
- Make consistent estimates from different surveys

To achieve these objectives, the set of auxiliary variables must be related to the topic of the survey. Calibration variables must be accessible with the timeliness and punctuality required and harmonized with those definitions in the EQLS. With these requirements, Core Social Variables[4] (or a subset of them) are the natural candidates for the set of calibrate variables. Core Social Variables provide an answer to the rapid evolution that social statistics have experienced in the last two decades at European level. This evolution has increasingly highlighted the need for a higher degree of integration of the concepts used across statistical instruments in order to define in a similar way statistical concepts present in various different surveys.

A limited number of core statistical variables are proposed for systematic introduction in all the EU social surveys. The objective is two-fold:

- To allow for better identification of specified populations across all the surveys, and a better description of those groups,
- To allow for socio-economic analysis based on the main structural variables

Table 2.1 shows availability of data in EQLS, LFS and EU SILC at country and NUTS2 level and the aggregation of variables.

[3] Larsen, M., Qing, S., Zhou, B. and Foulkes, M. (2011) Calibration Estimation and Longitudinal Survey Weights: Application to the NSF Survey of Doctorate Recipients. *Section on Survey Research Methods – JSM*.
Särndal, C. (2007) The calibration approach in survey theory and practice. *Component of Statistics Canada Catalogue no. 12-001.*
[4] A list of Core Social Variables was established by: Eurostat (2007) Task Force on Core Social Variables. Final report. *Eurostat. Methodologies and working papers*

Table 2.1. Availability of Core Social Variables[5] in Eurostat database by geographical breakdown.

	EQLS	EU-SILC		LFS	
	Variable	Country	NUTS2	Country	NUTS2
Demographic information					
Sex	YES	YES	NO	YES	NO
Age in completed years	YES	YES	NO	YES	NO
Country of birth	NO	NO	NO	NO	NO
Country of citizenship	YES	NO	NO	YES	NO
Legal marital status	NO	NO	NO	NO	NO
Consensual union	YES	NO	NO	NO	NO
Household size	YES	NO	NO	NO	NO
Household type	YES	YES	NO	NO	NO
Household economical activity	YES	NO	NO	NO	NO
Geographic information					
Country of residence	YES	YES	NO	YES	NO
Region of residence	YES	NO	NO	NO	NO
Degree of urbanisation	YES	YES	NO	YES	YES
Socio-economic information					
Self-declared activity/labour status	YES	NO	NO	NO	NO
Status in employment	YES	NO	NO	YES	NO
Occupation in employment	YES	YES	NO	YES	NO
Economic sector in employment	NO	NO	NO	YES	YES
Highest level of education attained	YES	YES	NO	YES	YES
Net monthly income of household	YES	NO	NO	NO	NO

Calibration with auxiliary information can improve the precision of estimates such as population totals and/or means and allows dealing with the common problems of non-response and coverage. Calibration has established itself as an important methodological instrument in large-scale production of statistics. With this approach quality of the auxiliary information is critical, and without appropriate auxiliary data reweighting is useless.

All the proposed calibration strategies in this project for EQLS will consider the calibration techniques introduced by Deville and Särndal (1992)[6]. These strategies will include – at least - a basic set of demographical variables based on the Eurostat's demographic statistics, namely:

a. Sex

b. Age groups defined considering EQLS sample sizes, in order to avoid distortions due to sample variability

c. NUTS2 regional level

[5] This table presents a complete list of the core social variables established by: Eurostat (2007*)* Task Force on Core Social Variables. Final report. *Eurostat. Methodologies and working papers*
[6] Deville, J.C. and Särndal, C.E. (1992) Calibration estimators in survey sampling. *JASA, vol 87 (418).*

Additional variables to be considered for calibration will be obtained from alternative sources.

2.2. Assessment of potential information sources for drawing calibration variables

The external sources considered in the project had to meet, at least, the following requirements:

- Availability for all EQLS-related countries
- Harmonized methodology across countries
- Certified quality

With these requirements, statistics of the European Statistical System arises as the relevant external sources to be taken into consideration. Because of their belonging to the same thematic aspect of social statistics or having common variables with the EQLS, the following sources are proposed:

- Eurostat's demographic statistics
- Labour Force Survey
- EU-SILC.

Since timeliness is a relevant quality dimension for the 3rd EQLS, other potential sources with a larger time lag between reference and publication dates or low frequency, such as censuses, are not recommended for the design of the weighting strategy of EQLS. However, they could be used for a retrospective reweighting in order to assess the quality of variables from surveys as calibration variables. Table 2.2 presents the main characteristics of each data source that have been considered in order to propose the additional variables for calibration.

Table 2.2. Key features of the information sources.

Source	Features
Eurostat's demographic statistics	• First option for auxiliary information. • Availability and periodicity depends on the variable considered. • As this source does not suffer from sampling error, data is in general more disaggregated than in survey-based sources.
Labour Force Survey	• Quarterly survey. • Data also available as annual means. • Largest European household sample survey (the quarterly survey in 2012 interviewed about 1.5 millions of individuals in EU27) • Non response around 20% in average • Covers 33 countries: the 28 Member States of the European Union, three EFTA countries (Iceland, Norway and Switzerland), and two EU candidate countries, the Former Yugoslav Republic of Macedonia and Turkey.
EU-SILC	• Yearly survey • Main source for the compilation of statistics on income, social inclusion and living conditions • Large sample size (minimum of 272.000 persons for EU27) • Non response around 20% in average • Coverage: the 28 Member States of the European Union and some other neighbouring countries, depending on the years.

2.2.1 Context of other survey practices: a case of the European Social Survey

The European Social Survey (ESocS) is a statistical operation whose weighting strategy could be compared with that of EQLS. As EQLS, ESocS is a biannual survey that is not included in the European Statistical System. ESocS covered 36 countries in its round 6 in 2012: the 28 Member States of the European Union, Albania, Iceland, Israel, Kosovo, Norway, Russian Federation, Switzerland and Turkey, with a sample size of 800 or 1.500 persons per country. The unit non-response rate is 33%.

Although the European Social Survey is not recommended as auxiliary information source - due to its sample size and non-response rate – an analysis of the weighting strategy of ESocS might be interesting as an example that can provide some contextual information on existing practice in the field of current surveys of European scale. ESocS methodology considers three different types of weights[7]:

- **Design weights** for each country, to correct for unequal probabilities for selection due to the sampling design used. Design weights are rescaled in a way that the sum of the final weights equals the country sampling size.
- **Calibration (post-stratification) weights.** These weights are computed using *age, gender, education and region* as calibration variables. The sources for these variables are the LFS for the 28 Member States.
- **Population size** weights to correct for population size when combining data from two or more countries.

It must be highlighted that the degree of urbanisation and the activity status are not considered into the set of calibration variables for this survey.

To produce weighted tables, design, calibration and population size weights should be combined according to the basic rules presented in the dissemination material of ESocS[8]. This practice differs with that of the dissemination of the weights of the 3rd EQLS, where different weights are provided for some aggregations of countries. The dissemination of population size weights could be a benchmark for Eurofound, since it facilitates the computation of aggregated weighted tables for any combination of countries and completes the dissemination of different weights in term of groups of countries.

2.2.2 Assessment of the calibration variables used in 3rd EQLS

For the sake of completeness, this section presents a quality assessment of the variables and sources actually applied for the calibration of 3rd EQLS, namely population by age and sex, households by household size and degree of urbanization. Tables 2.3 and 2.4 presents the definition of this variables and the sources that were used in the calibration of 3rd EQLS.

---

[7] ESS6 - 2012 Documentation Report. Edition 2.0. http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e02_0.pdf
[8] http://www.europeansocialsurvey.org/docs/methodology/ESS_weighting_data.pdf

Table 2.3 Calibration variables used in 3rd EQLS

Variable	Definition
Age and sex	• Sex coded by the interviewer. • Age provided by the respondent (last birthday assignment).
Region	• NUTS 2 level.
Household size	• Respondent's answer to the question "Including yourself, can you please tell me how many people live in this household?".
Urbanization	• Country-depending classifications based on different criteria (population size, level of urbanisation, etc.).

Table 2.4 Sources of the population marginal distribution used for calibration in 3rd EQLS by country (Source 3rd EQLS weighting report)

	Age by sex	Urbanization	Region	HH size
Austria	Eurostat 2011	Microcensus 2009	Microcensus 2009	Microcensus 2011
Belgium	Eurostat 2010	Orgassim 2011	Orgassim 2011	National register 2009
Bulgaria	Eurostat 2011	Microcensus 2009	Microcensus 2009	EU SILC 2010
Cyprus	Eurostat 2010	Projection 2002	Projection 2002	EU SILC 2010
Czech Republic	Eurostat 2011	NSI 2009	NSI 2009	EU SILC 2010
Germany	Eurostat 2011	Media analysis radio 2010	Media analysis radio 2010	Microcensus 2009
Denmark	Eurostat 2011	NSI 2011	NSI 2011	EU SILC 2010
Estonia	Eurostat 2011	NSI 2010	NSI 2010	HBS 2010
Greece	Eurostat 2011	Census 2011	Census 2011	EU SILC 2010
Spain	Eurostat 2011	NSI	NSI	HBS 2010
Finland	Eurostat 2011	NSI	NSI	National register 2010
France	Eurostat 2009	Census 2009	Census 2009	Census 2008
Hungary	Eurostat 2011	NSI 2010	NSI 2010	EU SILC 2010
Ireland	Eurostat 2011	Census 2006	Census 2006	EU SILC 2010
Italy	Eurostat 2011	NSI 2009	NSI 2009	EU SILC 2010
Lithuania	Eurostat 2011	NSI 2011	NSI 2011	EU SILC 2010
Luxembourg	Eurostat 2011	Census 2010	Census 2010	EU SILC 2010
Latvia	Eurostat 2011	Population register 2007	Population register 2007	EU SILC 2010
Malta	Eurostat 2011	-	-	HBS 2008
Netherlands	Eurostat 2011	NSI 2010	NSI 2010	National register 2008
Poland	Eurostat 2011	NSI 2010	NSI 2010	EU SILC 2010
Portugal	Eurostat 2011	Census 2001	Census 2001	EU SILC 2010
Romania	Eurostat 2010	NSI 2010	NSI 2010	EU SILC 2010
Sewden	Eurostat 2011	NSI 2009	NSI 2009	EU SILC 2010
Slovenia	Eurostat 2011	NSI 2009	NSI 2009	EU SILC 2010
Slovakia	Eurostat 2011	Census 2010	Census 2010	EU SILC 2010
UK	Eurostat 2011	NSI 2009	NSI 2009	EU SILC 2010

- **Age by sex.** As shown in Table 2.4, the statistical source used for calibration by age and sex in all countries was Eurostat's demographics statistics, the reference date being the latest available for each country. There is no other source of available at the European level able to provide this information to the same standard of consistency and completeness.

  The definition of the target universe of the EQLS, as people who have their usual residence in the country, fully complies with Eurostat's definition of population. The only minor lack of coherence is that derived of the time reference on which age is calculated. In the case of Eurostat, age is calculated as on 1 January of each year. However, age in EQLS is calculated from the last birthday and there will be a small positive difference in favour of EQLS estimation. This difference will be larger as the fieldwork gets closer to the end of the year. However, since the age groups used in the calibration are quite broad, in fact, ten years groups, this difference should not cause significant biases in the estimates. In any case, it is advisable that in later editions of EQLS, respondents were asked the year of birth in addition to or instead of age.

- **Household size.** The lack of availability of recent statistics of population by household size led to using household distribution by household size as a calibration variable in the original post-stratification of the 3rd EQLS. A reliable source for this distribution is EU SILC. The sample distribution of households by household size can be easily obtained for all countries, distinguishing between these cases where household selection is an actual sampling stage (most of the countries) and those cases where respondents are directly selected with no intermediate step. In the first case, the estimated distribution of households by size is straightforward, since each respondent is representative of a household. In the second case, an estimate of the number of households of size k can be obtained simply by dividing the number of people living in households of size k, by k. With this mechanism, household size becomes a variable of EQLS and therefore it can be used in common calibration processes.

  At this point one question arises: is it possible to compare the EQLS 2003 and 2007 results (calibrated by distribution of POPULATION by HH size) with those of EQLS 2011 (calibrated by distribution of HHs by HH size)?

  If no aggregation had been made, fitting EQLS to HHs by HH size distribution would have lead to approximately fit population by HH size.

  If x_i, X_i, y_i, Y_i, i=1…L are, respectively, the sampling and population totals of HHs by HH size (X) and persons by HH size (Y), h the number of HHs in the sample, H the number of HHs in the population, n the number of persons living in sampled HHs and N the number of persons in the population, then after calibration

  $$\frac{x_i}{h} = \frac{X_i}{H}, \text{ so } \frac{i \cdot x_i}{h} = \frac{i \cdot X_i}{H}$$

  Moreover, n/h is an estimator of N/H and then

  $$\frac{n}{h} \cong \frac{N}{H} \quad \text{and} \quad \frac{i \cdot x_i}{h \cdot \frac{n}{h}} \cong \frac{i \cdot X_i}{H \cdot \frac{N}{H}}, \quad \text{so} \quad \frac{y_i}{n} \cong \frac{Y_i}{N}$$

  hence, if no aggregation had been made, no recalibration would have been necessary in order to compare with 203 and 2007 EQLS. The only aggregation made has been actually made is that for “HHs of six or more persons” and this category only contains 2.1% of HHs (EU28), so recalibrate 1st and 2nd EQLS seems no necessary.

Table 2.4 shows a high level of harmonisation in the reference sources, EU SILC being the source used by 18 out of the 24 countries in EU27 where no national registers are available. According to the instructions on the 3rd EQLS questionnaire, household members are considered those whose habitual residence is the household [9], with specified exclusions and inclusions that are harmonised in both surveys. Therefore, definitions of both sources are consistent. According to the consultants, this harmonisation should be enhanced in next waves of EQLS, substituting at least the use of HBS (used in case of EE, ES, and MT in 2011) by data from EU-SILC for the sake of consistency of sources.

- **Region and urbanization level.** The reference marginal distributions for calibration of these two variables are obtained from different sources in the different Member States. For the sake of coherence and comparability, the harmonisation in these variables should be enhanced. Eurostat demographic statistics could be used as the source for NUTS2 population levels. As regards the sources for degree of urbanisation, though they often come from operations promoted by the National Statistics Offices, they are not homogeneous with regard to its methodology (micro census, population projections, etc.) and classification. DEGURBA classifications could become a good harmonised coding for the degree of urbanisation and LFS could provide with the information required to use such a harmonised variable in calibration. These issues are discussed in detail in both section 2.4 and chapter 4.

2.3. Methods of calibration

In general terms, the calibration of a sample consists of finding out a set of new weights (calibrated weights) close to original weights, which when applied to the sample, reproduce some predefined population indicators, which are considered known without or with small sampling error. The problem, thus stated, can be solved in mathematical terms, as a minimization problem (distance between original weights and calibrated weights) with restrictions (estimates to be met). In this framework, given a set of population restrictions, different methods of calibration consider different functions used to define the distance between the original and the calibrated weights. Specifically, two alternative weighting methodologies – iterative proportional fitting and generalised regression estimation[10] - will be considered for the definition of our weighting scenarios:

[9] Inclusions: People who normally live at the address but are away for less than 6 months; People away at work for whom this is the main address; Boarders and lodgers.
Exclusions: People aged 18+ who live elsewhere due to work; Spouses who are separated and no longer resident; People away for 6 months or more; People resident in country for less than 6 months.

[10] Deming, W.E. and Stephan, F.F. (1940) On least squares adjustment of sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics, 11*.
Deville, J.C. and Särndal, C.E. (1992) Calibration estimators in survey sampling. *JASA, vol 87 (418)*.
Deville, J.C., Särndal, C.E. and Sautory O. (1993) Generalized raking procedures in survey sampling. *JASA, vol 88 (423)*.

- **Iterative proportional fitting (IPF**), also known in finite population sampling as raking ratio. It is a method of fitting a multidimensional table to a given set of marginal population distributions. It consists, as the name suggests, of going sequentially adjusting the totals to each marginal and repeating the process until convergence. This method is equivalent to use as distance between the original weights $d_k$and the new weights w_k the following function:

$$\sum d_k G(\frac{w_k}{d_k})$$

  where $G(r) = r \log r - r + 1, r > 0$ and the sum is for all of the elements k of the sample.

- **Generalized Regression Estimation (GREG)** is a method of estimation in finite population sampling which improves direct estimators derived of the design, by fitting a regression model and using auxiliary information assumed known for the whole population. In the case of linear models, it can be shown that the estimates obtained by GREG can be expressed as weighted sums of calibrated weights that reproduce the totals of auxiliary variables. Therefore, besides to get improved estimates, the GREG is also used as a method of calibration. This method is equivalent to use the distance between the original weights $d_k$and the new weights w_k the following function:

$$\sum d_k G(\frac{w_k}{d_k})$$

  where $G(r) = \frac{1}{2}(r-1)^2, r \in \mathbb{R}$ and the sum is for all of the elements k of the sample.

  In those cases where the generalised regression model does not converge, this calibration methodology cannot be applied. Moreover, a calibration using GREG may generate negative weights that should be treated in the trimming phase. Although there no objective threshold in the percentage of negative weights above which the application of GREG should not be recommended, the interpretability of weighted estimates becomes more difficult as the proportion of negative weights becomes larger. None of these problems arises in the calibration exercise carried out in this report: the generalised regression models do always converge and - as presented in Table 3.1 - the percentage of negatives weights is lower than 1.5% for all GREG scenarios.

2.4. Proposed calibration scenarios

Calibration scenarios are defined as a combination of (1) the subset of calibration variables from the sources presented in subsection 2.2 and (2) each of the calibration methodologies described in 2.3. Some remarks and discussions should be made in order to support a proposal of such subsets of calibrations variables:

- *Education level*. This variable is completely harmonised among the different sources considered in subsection 2.2 (ISCED codification). Moreover, education is in general related to the non-response rate in social surveys. As a general trend, non-response rate is higher in those subpopulations with lower education levels, which may have additional difficulties to understand some questions and to feel comfortable during the interview process. These reasons lead to consider this variable for calibration.[11] ISCED codes will be aggregated in three main categories (ISCED 0-2; ISCED 3-4; ISCED 5-6) to guarantee that the samples sizes per category are not too small for a reliable calibration. The source for this variable for the calibration exercise is LFS, due to its larger sample size and the availability of information and NUTS2 level.

- *Activity status.* Since unemployed persons are - in general - easier to contact, labours status is one of the first candidates to be included in some of the weighting scenarios. As discussed in subsection 2.2, there are two main sources for this information in the ESS, namely LFS and EU-SILC. The main difference between both sources is the methodology to establish actual respondent’s status. In EU-SILC – as well as in EQLS – this status is self-assigned by the respondent, meanwhile in LFS the assignation is established through a series of objective rules. As shown in Table 2.5, these methodologies generate different estimations, specifically as regards with unemployment levels. For the sake of coherence in the definition of this variable, and regardless of the higher sampling sizes and more detailed geographic disaggregation of LFS, EU-SILC self-assigned activity status is preferable as a variable to be included in the calibration scenarios of EQLS. Moreover, to avoid potential seasonality issues, the annual statistic (rather than e.g. quarterly or so) on activity status will be considered.

Table 2.5. Activity/labour status according to EU-SILC and LFS (2012).

	EU-SILC			**LFS**		
	Employed	Unemployed	Inactive	Employed	Unemployed	Inactive
Belgium	50.2	7.9	41.5	49.2	4.0	46.8
Bulgaria	48.5	11.3	39.8	46.6	6.5	46.9
Czech Republic	53.7	5.2	40.7	54.5	4.1	41.4
Denmark	55.0	3.8	40.6	58.4	4.8	36.8
Germany	55.0	4.6	40.3	56.8	3.3	39.9
Estonia	54.8	6.1	38.6	55.2	6.1	38.7
Ireland	Not available[12]			51.1	8.8	40.1
Greece	41.5	12.8	45.5	40.1	12.8	47.0
Spain	46.3	15.2	38.3	44.5	14.8	40.7
France	52.9	5.2	41.5	51.1	5.6	43.3
Croatia	39.0	14.8	45.8	38.1	7.2	54.7
Italy	45.5	7.2	47.2	44.0	5.3	50.7
Cyprus	57.4	6.2	35.9	55.9	7.5	36.6
Latvia	49.7	10.9	38.9	50.7	9.0	40.3
Lithuania	49.9	8.7	40.4	50.0	7.7	42.3
Luxembourg	57.0	3.3	39.3	55.7	3.0	41.2
Hungary	46.8	6.9	45.8	46.3	5.7	48.0
Malta	48.9	3.0	48.0	48.2	3.3	48.5
Netherlands	57.1	5.2	37.5	61.8	3.4	34.8
Austria	54.7	3.3	41.8	58.8	2.7	38.5
Poland	52.1	6.6	40.9	50.2	5.6	44.1

[11] As pointed out in the Quality Assessment report of the 3rd EQLS, the distribution of groups by ‘education’ in certain countries has noteable differences between the EQLS simple and population statistics. For instance, in the UK, the population with low education (ISCED levels 0–2) was oversampled and population with intermediate attainment levels (ISCED 3–4) was undersampled. Certain other differences stand out in Spain, Lithuania and some other countries.

[12] EU-SILC 2011 was applied in the calibration for Ireland.

	EU-SILC			LFS		
	Employed	Unemployed	Inactive	Employed	Unemployed	Inactive
Portugal	50.2	9.5	40.2	51.4	9.5	39.0
Romania	52.5	2.2	45.2	51.1	3.9	45.1
Slovenia	51.5	6.9	41.6	52.5	5.1	42.5
Slovakia	51.9	7.9	39.7	50.9	8.3	40.8
Finland	51.6	5.7	41.8	55.2	4.6	40.3
Sweden	60.9	3.0	35.8	58.8	5.1	36.1
United Kingdom	60.6	3.4	35.9	57.7	4.9	37.3
Iceland	69.2	5.2	24.8	75.1	4.8	20.1
Norway	63.4	1.7	34.6	69.2	2.2	28.6
Switzerland	64.4	1.4	33.9	65.3	2.9	31.8

- *Citizenship*. Sample sizes of non-nationals (see Table 2.6) are generally too small to safely use it as calibration variable. In addition, the heterogeneity of the non-national subpopulations in terms of country of origin and level of mastering of the national language may induce some biases in the non-national population participating in the survey. The use of citizenship for calibration could tend to increase these potential biases. For these reasons, citizenship is not finally considered in the definition of the weighting scenarios.

Table 2.6. Number and proportion of non-national respondents in the 3rd EQLS sample (unweighted).

	Number	%		Number	%
Austria	45	4.4	Latvia	33	3.3
Belgium	51	5.0	Malta	1	0.1
Bulgaria	18	1.8	Netherlands	29	2.9
Cyprus	37	3.7	Poland	2	0.1
Czech Republic	9	0.9	Portugal	45	4.4
Germany	201	6.6	Romania	4	0.3
Denmark	30	2.9	Sweden	25	2.5
Estonia	157	15.7	Slovenia	3	0.3
Greece	78	7.8	Slovakia	7	0.7
Spain	122	8.1	UK	118	5.2
Finland	6	0.6	Turkey	39	1.9
France	107	4.7	Croatia	3	0.3
Hungary	1	0.1	FYROM	9	0.9
Ireland	103	9.8	Kosovo	35	3.3
Italy	33	1.5	Serbia	4	0.4
Lithuania	7	0.6	Montenegro	47	4.7
Luxembourg	364	36.2	Iceland	0	0.0

- *Degree of urbanisation*. This variable could be an obvious candidate to be used in the calibration procedure. However, some modification in the geocoding of EQLS is required before its application. For the 4th EQLS wave the consultant recommends to include as stratification variable EUROSTAT's DEGURBA codes for the very beginning of the sampling design. Specifically, strata should be defined using all the three codification levels: densely populated area (Code 1), intermediate density area (Code 2) and thinly populated area (Code 3).

A cross-cutting issue that have an impact in all the variables is the geographical breakdown to be considered for the computation of the weights. At this point, there is a clear trade off in using the geographically detailed disaggregation: the reference subpopulations tend to be more homogeneous as they get smaller but, at the same time, the corresponding subsamples tend to be also small. The reduction of the subsample size may have an impact in both (1) the convergence of the calibration algorithms and (2) the reliability of the calibration itself, since the features of a very small subsample are actually projected to the whole subpopulation in the calibration procedure.

In the definition of the calibration scenarios, this discussion translates into the selection of NUTS 2, NUTS 1 or Country Level as the geographical breakdown for the calibration variables. At this point, the following criteria are applied in all the weighting scenarios:

- The sample size of EQLS allows the usage of a breakdown at NUTS 2 level when no other breakdown is considered. However, when additional features are considered (such as gender by age classification), the subsamples in each subpopulation (i.e. gender by age by NUTS 2 or gender by age by NUTS 1) are two small of even zero. This fact generates relevant convergence problems in the computation algorithms and, when converging, the calibration weights produced by these algorithms cannot be considered as reliable. For these reasons, the six weighting scenarios consider (1) total population at NUTS 2 level and (2) sex by age population at country (NUTS 0) level. NUTS 1 level is not considered since sex by age subsamples are also small at this level.
- The new calibration variable Education level is obtained from LFS at NUTS 2 level.
- Finally, Activity status is obtained from EU-SILC for the coherence reasons discussed above. Since EU-SILC does not disseminate this information at NUTS 2 level, activity status is considered at country level for the calibration procedures.

Table 2.7 shows the six proposed scenarios. All six scenarios will contain the set of variables used for the original calibration of EQLS (number of households by household size at national level, as well as population by age cross with gender) and a different set of additional variables (Education level, Activity status or both).

Table 2.7. Proposal of calibration scenarios.

	Additional Variables	Source	Level	Calibration Method
Scenario 1	Education level Activity status	LFS EU-SILC	NUTS 2 Country level	IPF
Scenario 2	Education level	LFS	NUTS2	IPF
Scenario 3	Activity status	EU-SILC	Country level	IPF
Scenario 4	Education level Activity status	LFS EU-SILC	NUTS 2 Country level	GREG
Scenario 5	Education level	LFS	NUTS 2	GREG
Scenario 6	Activity status	EU-SILC	Country level	GREG

Chapter 3: Implementation and comparison of the weighting scenarios

3.1. Implementation of the weighting scenarios

The computations of the six weights corresponding to the scenarios in Table 2.5 has been implemented using the statistical software R. Specifically, the Consultant has applied functions from the R Cran packages[13]:

- *Foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, etc.* This package has provided functions for reading and writing data stored by some versions of Epi Info, Minitab, S, SAS, SPSS, Stata, Systat and Weka and for reading and writing some dBase files.
- *Car: Companion to Applied Regression.* This package accompanies J. Fox and S. Weisberg, An R Companion to Applied Regression, Second Edition, Sage, 2011.
- *Plyr: Tools for splitting, applying and combining data.* Plyr is a set of tools that solves a common set of problems, such as to break a big problem down into manageable pieces, operate on each pieces and then put all the pieces back together.
- *Survey: Analysis of complex survey samples.* This package has been applied to estimate variances by Taylor series linearisation as well as for calibration.

The complete R code to compute the weights is presented as a companion file, with the corresponding documentation to facilitate the replication of the computations and its application to future waves of EQLS.

3.2. Criteria to compare weighting scenarios

The selection of the most suitable weighting scenario is a multicriteria choice that should consider the trade-off between bias reduction and increment of variance intrinsic to any calibration procedure. In this analysis four complementary criteria to evaluate the weighting scenarios are considered:

a) Level of distortion of the design weights

b) Impact on the sampling errors

c) Level of bias in those variables for which official data are available and are not included in the calibration procedure

d) Level of distortion of the estimations obtained with the current calibration methodology of EQLS.

A full description of the indicators used in the evaluation of the scenarios is presented in the Table below:

[13] Foreign, Car, Plyr, Survey

Table 3.1. Indicators for the comparison of the alternative weighting scenarios.

Indicator	Formula	Interpretation
Distortion of weights	$\frac{\mathbf{weight_{i2}}}{\mathbf{weght_{i1}}}$ Where $weight_{i1}$ and $weight_{i2}$ are two weights corresponding to the same unit i but generated by different process, for example, calibrated weights and design weights	Used to compare two sets of weights. When the ratio is one weights are the same; in other cases, its size is a measure of the distortion: ratios greater (lower) than one mean $weights_2$ are greater (lower) than $weights_1$
Coefficient of variation	$\boldsymbol{CV} = \frac{\sqrt{\boldsymbol{V}\left(\hat{\hat{\boldsymbol{Y}}}\right)}}{\hat{\hat{\boldsymbol{Y}}}} \cdot \mathbf{100}$ Where $\boldsymbol{V}\left(\hat{\hat{\boldsymbol{Y}}}\right)$ is the sampling variance of the estimation $\hat{\hat{\boldsymbol{Y}}}$.	As indicators of sampling error, coefficients of variation (as percentage) have been computed. The coefficient of variation (CV) is the ratio of the standard deviation to the mean, and as a result, it is la relative measure of precision. It can take only positive values and, as it does not depend on the unit of measurement, it can be calculated for all questions jointly. The lower the CV the smaller the relative sampling error. It can take only positive values.
Bias or difference between estimates	$\mathbf{D_\infty(X, Y) = max_i\{\lvert x_i - y_i\rvert\}}$ Where x_i and y_i are estimates from different statistical operations or from the same operation but with different sampling weights, of the same population characteristic with several dimensions i=1 to k (for example, the different proportions of people with each level of education). And the max is computed over the i index	This is known as the infinity norm and it is used to measure the distance between to vectors as the maximum distance between each of its dimensions. The interpretation is very straightforward, it takes only positive or zero values where zero corresponds to equality and the greater the norm, the greater the difference between the estimates. As with every norm that tries to measure the difference between two vectors by means of one only quantity, some information is lost but, in the case of proportions, this is not too important, because their components are bounded and the norm is bounded itself ranging from 0 to 1 (or 0 to 100 with percentages).

Ad. A) Level of distortion of the design weights

The first criterion considered is the distortion of the design weight caused by the process of reweighting. The calibration process is equivalent to an optimization problem: there is an objective to minimize (a function of distance between the original design and the calibrated weights) provided a given marginal distributions of the calibration variables. In this framework, the criterion of selecting as the most suitable scenario that with the lowest weight distortion arises in a natural way.

There are different types of distortion to be considered. First of all, an analysis of the distribution of the ratios between the original design and the calibrated weights should be implemented. Moreover, since calibration with GREG[14] can distort the design weights to generate negative values, the number and percentages of cases with negative weights should be compared. Finally, an analysis of the extreme weights (both large and small) is required for a discussion on the trimming strategy in Section 4.

[14] Only those scenarios using GREG as calibration technique can generate negative weights. The Iterative Proportional Fitting technique implemented in R always produces non-negative weights.

Analysis of the ratio of the calibrated weight over the design weight

The closest approximation to the design weight in 3rd EQLS is variable w_1 in the survey database elevated to the total population aged 18 or older in the country (instead of the total sample as in EQLS). So a new weight w_0 was defined as

$$w_0 = w_1 \frac{w_3}{\sum w_1}.$$

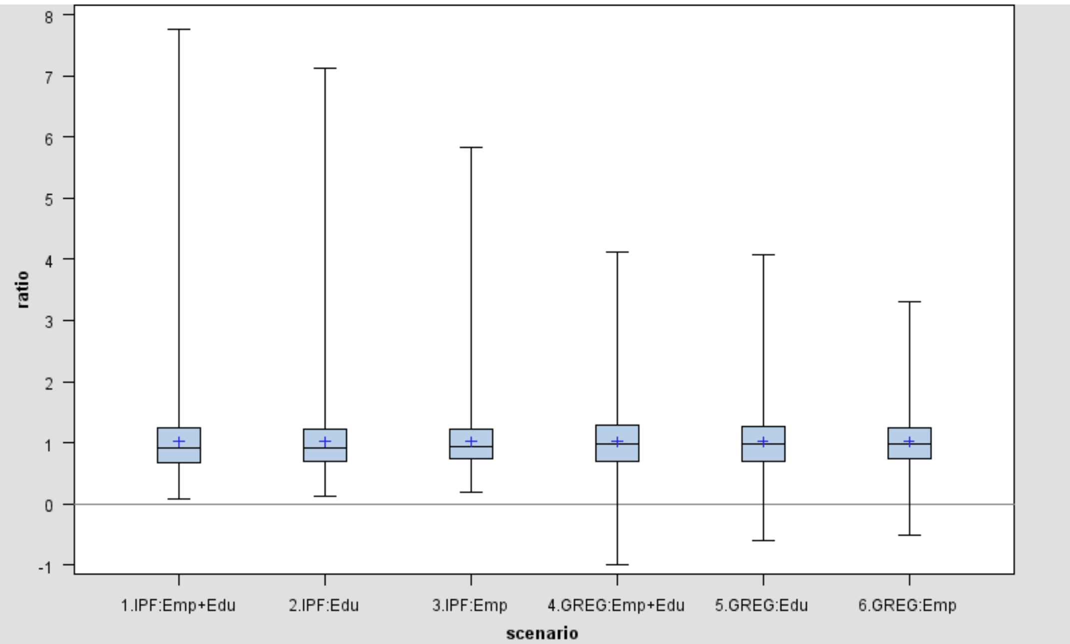
where w_1 and w_3 are the weights included in EQLS database, and the sum extends over each country (w_1 is the household weight, and w_3 the country population size). The definition of this new weight is required for the analysis, since in the calibration procedure, marginal distributions are related to subpopulation totals, not to sampling totals.

An aggregated descriptive analysis of the ratios of calibrated weights (w_{GREG} or w_{IPF}) and w_0 was performed for each scenario with the application of the following formula:

$$\text{Level of distortion} = (w_{GREG} \text{ or } w_{IPF})/w_0$$

The Box-Plot chart in Figure 3.1 shows the distribution of these ratios in the six scenarios weights. There are no major differences between the six scenarios considering the central box of the chart (including from the first quartile to the third, containing 50% of the ratios). However, the remaining 50% are more dispersed in the scenarios using IPF. The figures in Table 3.2 confirm this conclusion. Differences between quartiles are quite small, but for the percentiles lower 0.05 (P05) and upper 0.05 (P95), as well as for maximum and minimum major discrepancies are observed: the scenarios IPF exhibits greater dispersion that those based in GREG (even, as already discussed, the latter methodology may generate some negative values).

Figure 3.1. Distribution of the ratios of calibrated and design weights (EU27)



Statistics by scenario						
Min	0.083981	0.132353	0.189657	-0.99428	-0.59181	-0.50767
Mean	1.022472	1.02199	1.022437	1.022234	1.022087	1.022358
Std Dev	0.550891	0.524539	0.422628	0.514016	0.492755	0.403489
Max	7.753567	7.118146	5.827795	4.127582	4.084355	3.299643

Table 3.2. Percentiles, minimum and maximum of the ratios of calibrated and design weights (EU27).

	Scenario	Min	P5	P10	P25	P50	P75	P90	P95	Max
IPF	Scenario 1	0.08	0.38	0.49	**0.66**	0.91	1.24	**1.67**	**2.03**	**7.75**
	Scenario 2	0.13	0.41	0.51	0.69	0.92	1.23	1.64	1.97	7.12
	Scenario 3	0.19	0.51	0.59	0.74	0.94	1.22	1.54	1.80	5.83
GREG	Scenario 4	**-0.99**	**0.27**	**0.45**	0.69	0.98	**1.30**	1.66	1.92	4.13
	Scenario 5	-0.59	0.32	0.48	0.70	0.98	1.27	1.62	1.89	4.08
	Scenario 6	-0.51	0.44	0.56	0.75	0.98	1.25	1.53	1.78	3.30

As can be seen in Table 3.3, when weight ratios are analysed by country the dispersion is always higher in those scenarios which use IPF. Moreover, only 7 of 27 countries have ratios greater than 5 (which occurs only in the three IPF scenarios) being UK, Poland and Denmark the countries with higher ratios. Figures 3.2 to 3.4 suggest that in these three countries with larger distortion, the use of GREG-based scenarios is even more appropriate.

Figure 3.2. Distribution of the ratios of calibrated and design weights (United Kingdom).

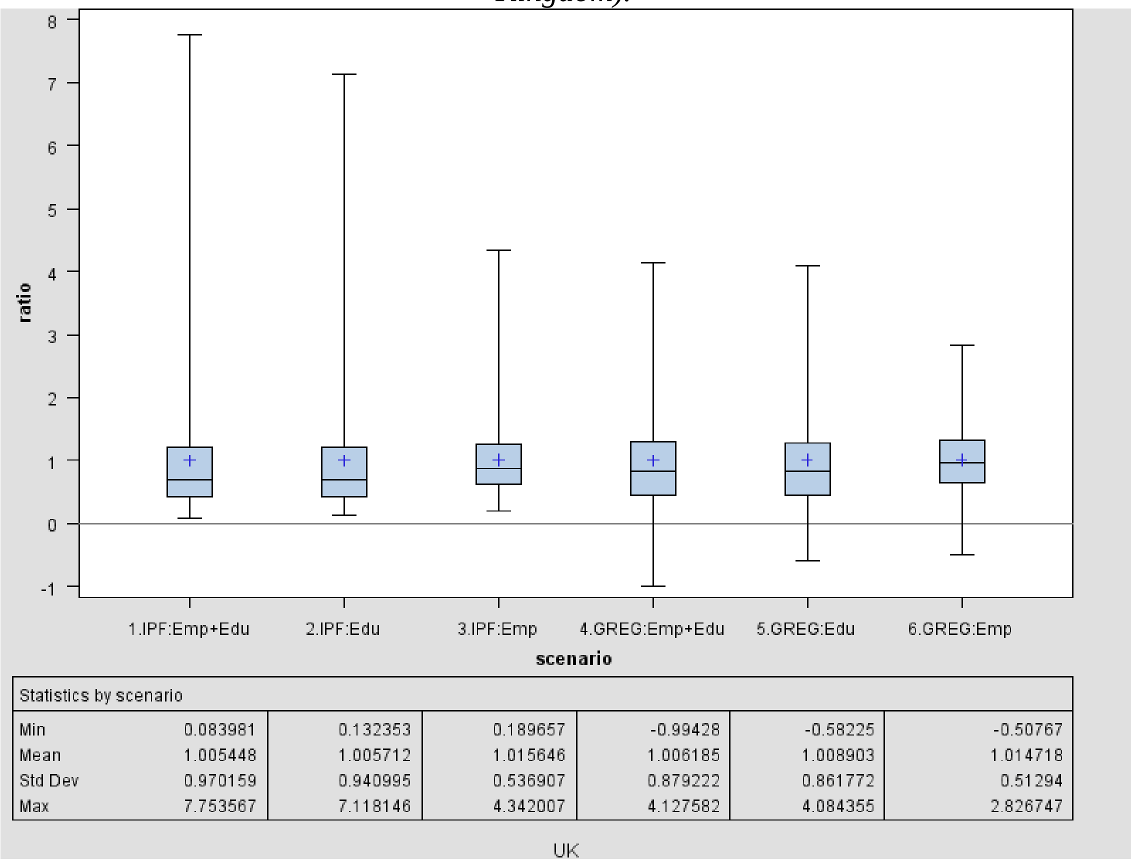
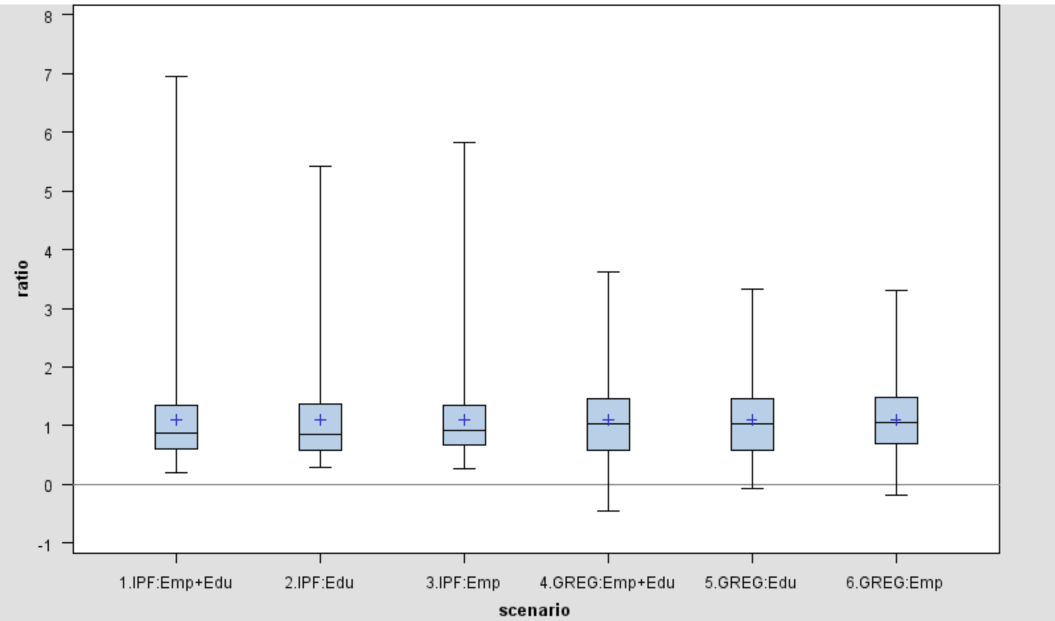


Figure 3.3. Distribution of the ratios of calibrated and design weights (Denmark).

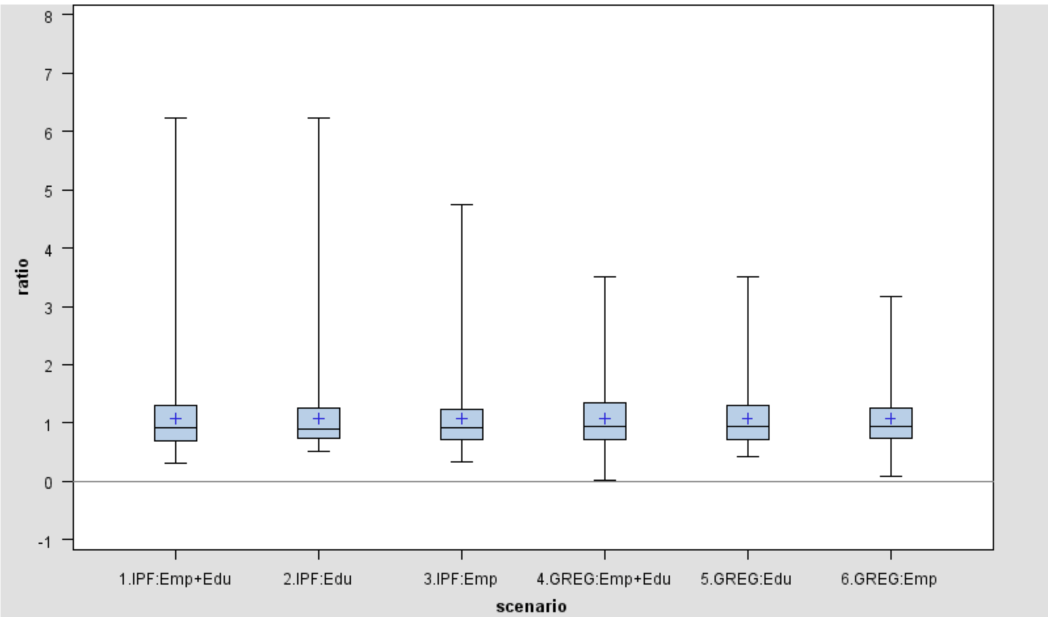


Statistics by scenario

	1.IPF:Emp+Edu	2.IPF:Edu	3.IPF:Emp	4.GREG:Emp+Edu	5.GREG:Edu	6.GREG:Emp
Min	0.195753	0.284671	0.276398	-0.4468	-0.06852	-0.17087
Mean	1.100863	1.100604	1.101066	1.101051	1.101363	1.101236
Std Dev	0.780069	0.750536	0.663648	0.684502	0.670498	0.588781
Max	6.956347	5.425713	5.827795	3.626609	3.332195	3.299643

Denmark

Figure 3.4. Distribution of the ratios of calibrated and design weights (Poland).



Statistics by scenario

	1.IPF:Emp+Edu	2.IPF:Edu	3.IPF:Emp	4.GREG:Emp+Edu	5.GREG:Edu	6.GREG:Emp
Min	0.303188	0.519628	0.331045	0.014686	0.416085	0.079108
Mean	1.08334	1.082821	1.082191	1.082235	1.081995	1.081885
Std Dev	0.564081	0.540167	0.524341	0.535756	0.514871	0.50615
Max	6.223873	6.23543	4.734975	3.507706	3.50751	3.164495

Poland

Table 3.3. Percentiles, minimum and maximum of the ratios of calibrated and design weights by country.

	Scenario 1				Scenario 2				Scenario 3				Scenario 4				Scenario 5				Scenario 6			
	Min	P25	P75	Max	Min	P25	P75	Max	Min	P25	P75	Max	Min	P25	P75	Max	Min	P25	P75	Max	Min	P25	P75	Max
Austria	0.20	0.55	1.28	5.05	0.21	0.56	1.29	5.21	0.29	0.67	1.33	3.17	-0.40	0.58	1.47	3.02	-0.35	0.56	1.45	3.05	-0.08	0.68	1.38	2.39
Belgium	0.37	0.80	1.32	2.64	0.47	0.83	1.26	2.04	0.44	0.82	1.28	2.46	0.10	0.83	1.33	2.14	0.33	0.85	1.28	1.84	0.27	0.85	1.32	2.10
Bulgaria	0.38	0.70	1.10	5.35	0.37	0.70	1.10	4.65	0.39	0.76	1.01	3.46	0.10	0.69	1.16	3.24	0.12	0.69	1.18	3.04	0.08	0.75	1.04	2.71
Cyprus	0.23	0.70	1.39	3.34	0.46	0.67	1.50	2.92	0.25	0.69	1.47	3.17	-0.19	0.71	1.51	2.70	0.37	0.69	1.48	2.54	-0.08	0.74	1.48	2.63
Czech Republic	0.14	0.55	1.29	5.74	0.20	0.57	1.26	5.61	0.41	0.74	1.21	4.18	-0.45	0.64	1.36	3.69	-0.18	0.69	1.30	3.68	0.20	0.74	1.24	2.88
Germany	0.12	0.63	1.34	4.47	0.13	0.64	1.33	3.57	0.30	0.74	1.27	3.24	-0.66	0.72	1.39	2.65	-0.59	0.73	1.37	2.51	0.00	0.76	1.33	2.36
Denmark	0.20	0.60	1.35	6.96	0.28	0.59	1.38	5.43	0.28	0.68	1.36	5.83	-0.45	0.57	1.47	3.63	-0.07	0.58	1.47	3.33	-0.17	0.70	1.49	3.30
Estonia	0.19	0.64	1.24	3.49	0.23	0.66	1.27	3.47	0.37	0.78	1.23	2.48	-0.41	0.70	1.34	2.53	-0.17	0.70	1.37	2.51	0.11	0.78	1.27	2.11
Greece	0.39	0.70	1.09	2.23	0.38	0.69	1.09	2.23	0.47	0.72	1.08	2.02	0.22	0.70	1.12	1.86	0.20	0.69	1.13	1.86	0.38	0.72	1.11	1.76
Spain	0.24	0.63	1.28	3.50	0.26	0.62	1.29	3.16	0.49	0.84	1.15	2.42	-0.26	0.70	1.33	2.52	-0.16	0.70	1.33	2.38	0.35	0.85	1.17	2.04
Finland	0.19	0.44	1.44	5.58	0.18	0.45	1.44	5.73	0.25	0.45	1.53	4.23	-0.29	0.45	1.59	3.22	-0.30	0.46	1.59	3.24	-0.04	0.48	1.63	3.03
France	0.35	0.71	1.33	4.08	0.41	0.72	1.33	3.34	0.46	0.86	1.23	1.88	0.09	0.71	1.37	2.54	0.16	0.70	1.36	2.35	0.32	0.87	1.24	1.69
Hungary	0.32	0.79	1.19	2.37	0.37	0.80	1.16	1.74	0.36	0.84	1.18	1.97	0.01	0.81	1.20	1.95	0.16	0.82	1.18	1.62	0.09	0.86	1.18	1.73
Ireland	0.52	0.77	1.19	2.33	0.54	0.78	1.18	2.28	0.53	0.78	1.16	2.20	0.40	0.77	1.24	2.02	0.41	0.78	1.22	2.00	0.40	0.77	1.19	1.97
Italy	0.33	0.66	1.28	4.64	0.34	0.67	1.28	4.65	0.42	0.70	1.29	4.10	0.02	0.68	1.34	3.30	0.08	0.69	1.33	3.30	0.27	0.71	1.33	3.01
Lithuania	0.28	0.60	1.20	2.70	0.30	0.60	1.16	2.84	0.47	0.73	1.14	2.10	-0.05	0.58	1.27	2.15	0.05	0.56	1.25	2.15	0.32	0.74	1.21	1.82
Luxembourg	0.54	0.77	1.28	2.68	0.62	0.74	1.27	2.56	0.56	0.78	1.31	2.71	0.44	0.76	1.35	2.27	0.56	0.72	1.34	2.22	0.46	0.77	1.41	2.29
Latvia	0.29	0.69	1.13	2.35	0.30	0.68	1.13	2.14	0.31	0.70	1.08	2.33	0.05	0.70	1.16	1.91	0.08	0.68	1.17	1.87	0.11	0.69	1.14	1.95
Malta	0.35	0.72	1.10	3.25	0.34	0.70	1.10	3.26	0.60	0.87	1.13	1.63	0.11	0.70	1.11	2.68	0.09	0.70	1.11	2.68	0.53	0.89	1.13	1.54
Netherlands	0.31	0.78	1.23	4.20	0.35	0.80	1.22	2.41	0.36	0.80	1.25	3.51	-0.01	0.79	1.27	2.86	0.09	0.80	1.26	2.02	0.13	0.81	1.25	2.62
Poland	0.30	0.70	1.30	6.22	0.52	0.73	1.26	6.24	0.33	0.73	1.24	4.73	0.01	0.71	1.34	3.51	0.42	0.72	1.30	3.51	0.08	0.74	1.26	3.16
Portugal	0.46	0.67	1.09	2.57	0.49	0.70	1.13	2.43	0.46	0.70	1.10	2.74	0.31	0.67	1.18	2.08	0.39	0.70	1.17	2.01	0.31	0.68	1.17	2.14
Romania	0.18	0.61	1.19	3.04	0.36	0.74	1.11	2.02	0.41	0.65	1.15	1.71	-0.46	0.64	1.22	2.13	0.14	0.74	1.12	1.72	0.26	0.67	1.18	1.57
Sweden	0.26	0.65	1.26	3.47	0.38	0.67	1.29	3.00	0.32	0.74	1.25	2.53	-0.12	0.65	1.34	2.51	0.20	0.66	1.34	2.37	0.05	0.75	1.28	2.14
Slovenia	0.45	0.79	1.16	1.81	0.64	0.86	1.12	1.64	0.47	0.82	1.17	1.50	0.31	0.81	1.18	1.66	0.59	0.86	1.14	1.55	0.37	0.81	1.19	1.47
Slovakia	0.37	0.67	1.23	2.64	0.37	0.69	1.26	2.64	0.35	0.68	1.20	2.70	0.14	0.69	1.27	2.18	0.17	0.73	1.29	2.11	0.05	0.71	1.26	2.21
UK	0.08	0.41	1.21	7.75	0.13	0.42	1.22	7.12	0.19	0.62	1.27	4.34	-0.99	0.44	1.31	4.13	-0.58	0.45	1.28	4.08	-0.51	0.65	1.33	2.83

The main conclusion according to the first comparison criterion is that GREG based scenarios seem more appropriate due to the smaller distortion of design weights and the small number of negative weights. Among GREG scenarios 4 to 6, scenario 6 exhibits the lowest dispersion and fewer negative weights.

Analysis of negative weights (GREG-based calibration scenarios)

The interpretation and application of negative weights could be difficult and may generate potential errors by unskilled users. Moreover, if the number of negative weights were very high, some subpopulation estimate could be higher than that of the total population. For these reasons, negatives weights need to be corrected in the trimming phase (see Section 4) and those scenarios with a large prevalence of negative weights should be excluded from EQLS weighting strategy.

As can be seen in Table 3.4, the percentage of negative weights in the three GREG scenarios is quite small, about 1% (in the worst case, scenario 4. only 508 records have negative weights). Furthermore, albeit the number of countries with negative weights is respectively 13, 8 and 5 for scenarios 4, 5 and 6, only two countries (Finland and United Kingdom) have a percentage of cases with negative weights over 5% on scenarios 4 and 5 (in the scenario 6 no country surpasses even 2% of negative weights). As a consequence of this small number of negative weights, no scenario should be discarded if only negative weights criterion is considered.

Table 3.4. Number and percentage of cases with negative weights in those scenarios using GREG (EU27).

	Scenario 4		Scenario 5		Scenario 6	
	Negative weights	%	Negative weights	%	Negative weights	%
ALL	**508**	**1.5%**	**341**	**1.0%**	**54**	**0.2%**
Austria	21	2.1%	23	2.3%	4	0.4%
Belgium	0	0.0%	0	0.0%	0	0.0%
Bulgaria	0	0.0%	0	0.0%	0	0.0%
Cyprus	9	0.9%	0	0.0%	5	0.5%
Czech Republic	39	4.0%	15	1.5%	0	0.0%
Germany	87	2.9%	65	2.2%	0	0.0%
Denmark	25	2.5%	21	2.1%	3	0.3%
Estonia	23	2.3%	29	3.0%	0	0.0%
Greece	0	0.0%	0	0.0%	0	0.0%
Spain	21	1.4%	14	0.9%	0	0.0%
Finland	66	6.9%	58	6.0%	11	1.1%
France	0	0.0%	0	0.0%	0	0.0%
Hungary	0	0.0%	0	0.0%	0	0.0%
Ireland	0	0.0%	0	0.0%	0	0.0%
Italy	0	0.0%	0	0.0%	0	0.0%
Lithuania	4	0.4%	0	0.0%	0	0.0%
Luxembourg	0	0.0%	0	0.0%	0	0.0%
Latvia	0	0.0%	0	0.0%	0	0.0%
Malta	0	0.0%	0	0.0%	0	0.0%
Netherlands	1	0.1%	0	0.0%	0	0.0%
Poland	0	0.0%	0	0.0%	0	0.0%
Portugal	0	0.0%	0	0.0%	0	0.0%

	Scenario 4		Scenario 5		Scenario 6	
	Negative weights	%	Negative weights	%	Negative weights	%
Romania	29	1.9%	0	0.0%	0	0.0%
Sweden	3	0.3%	0	0.0%	0	0.0%
Slovenia	0	0.0%	0	0.0%	0	0.0%
Slovakia	0	0.0%	0	0.0%	0	0.0%
UK	180	8.1%	116	5.4%	31	1.4%

Analysis of extreme weights

Table 3.5 presents the percentage of extreme weights per country. Although the definition of 'extreme' is ad hoc, the table presents the prevalence of those weights that are larger than 3 and 4 or smaller than 0.3 and 0.4. Notice that the percentage of cases smaller than a given value does also include the negative weights.

An analysis of this tables shows that the prevalence of extremely high weights is not relevant in any of the six scenarios, with the percentage of weights larger than 3 being lower than 1%. Extremely large weights are more frequent in Denmark, Finland and UK, but they appear in less than 5% of the cases. Since one of the main dangers of calibration is the potential generation of extremely high weights that may fictitiously increase the representativeness of potential outliers, this conclusion is quite positive and supports the robustness of the sampling methodology of EQLS.

As regards with extreme small weights, their prevalence is higher, specifically in GREG-based scenarios. This fact suggests the convenience of the application of a strategy to treat extreme weights. This issue is discussed in depth in Section 4.

Table 3.5. Percentage of cases with extreme weights (EU27).

	Scenario	<0.3	<0.4	>3	>4		Scenario	<0.3	<0.4	>3	>4
Total	1	2.2%	5.6%	0.9%	0.3%	Ireland	1	0.0%	0.0%	0.0%	0.0%
	2	1.7%	4.7%	0.7%	0.2%		2	0.0%	0.0%	0.0%	0.0%
	3	0.3%	1.5%	0.3%	0.0%		3	0.0%	0.0%	0.0%	0.0%
	4	5.5%	8.3%	0.4%	0.0%		4	0.0%	0.1%	0.0%	0.0%
	5	4.6%	7.1%	0.4%	0.0%		5	0.0%	0.0%	0.0%	0.0%
	6	2.1%	3.9%	0.0%	0.0%		6	0.0%	0.0%	0.0%	0.0%
Austria	1	1.6%	10.0%	2.1%	0.3%	Italy	1	0.0%	2.5%	0.6%	0.1%
	2	1.8%	9.5%	2.2%	0.2%		2	0.0%	2.2%	0.7%	0.1%
	3	0.4%	2.4%	0.2%	0.0%		3	0.0%	0.0%	0.4%	0.1%
	4	10.8%	17.7%	0.1%	0.0%		4	3.8%	7.8%	0.3%	0.0%
	5	10.9%	17.7%	0.1%	0.0%		5	3.2%	7.7%	0.3%	0.0%
	6	5.5%	9.4%	0.0%	0.0%		6	0.2%	2.2%	0.1%	0.0%
Belgium	1	0.0%	0.4%	0.0%	0.0%	Lithuania	1	1.8%	5.9%	0.0%	0.0%
	2	0.0%	0.0%	0.0%	0.0%		2	0.0%	4.9%	0.0%	0.0%
	3	0.0%	0.0%	0.0%	0.0%		3	0.0%	0.0%	0.0%	0.0%
	4	1.5%	2.9%	0.0%	0.0%		4	6.1%	10.5%	0.0%	0.0%
	5	0.0%	1.4%	0.0%	0.0%		5	4.5%	10.0%	0.0%	0.0%
	6	0.4%	1.8%	0.0%	0.0%		6	0.0%	0.4%	0.0%	0.0%

	Scenario	<0.3	<0.4	>3	>4		Scenario	<0.3	<0.4	>3	>4
Bulgaria	1	0.0%	0.3%	0.3%	0.1%	Luxembourg	1	0.0%	0.0%	0.0%	0.0%
	2	0.0%	0.3%	0.2%	0.1%		2	0.0%	0.0%	0.0%	0.0%
	3	0.0%	0.1%	0.3%	0.0%		3	0.0%	0.0%	0.0%	0.0%
	4	1.4%	3.9%	0.1%	0.0%		4	0.0%	0.0%	0.0%	0.0%
	5	0.7%	2.7%	0.1%	0.0%		5	0.0%	0.0%	0.0%	0.0%
	6	0.6%	1.1%	0.0%	0.0%		6	0.0%	0.0%	0.0%	0.0%
Cyprus	1	2.2%	4.7%	0.5%	0.0%	Latvia	1	0.2%	3.7%	0.0%	0.0%
	2	0.0%	0.0%	0.0%	0.0%		2	0.1%	2.8%	0.0%	0.0%
	3	3.1%	4.1%	1.5%	0.0%		3	0.0%	3.8%	0.0%	0.0%
	4	5.6%	10.5%	0.0%	0.0%		4	3.2%	5.9%	0.0%	0.0%
	5	0.0%	1.7%	0.0%	0.0%		5	2.5%	5.3%	0.0%	0.0%
	6	6.1%	13.2%	0.0%	0.0%		6	3.8%	4.5%	0.0%	0.0%
Czech Republic	1	9.0%	17.2%	1.9%	0.4%	Malta	1	0.0%	0.3%	0.2%	0.0%
	2	7.7%	18.2%	1.5%	0.4%		2	0.0%	0.3%	0.1%	0.0%
	3	0.0%	0.0%	0.2%	0.2%		3	0.0%	0.0%	0.0%	0.0%
	4	11.5%	15.3%	1.1%	0.0%		4	1.5%	3.2%	0.0%	0.0%
	5	12.0%	16.7%	1.0%	0.0%		5	1.4%	3.6%	0.0%	0.0%
	6	0.6%	2.6%	0.0%	0.0%		6	0.0%	0.0%	0.0%	0.0%
Germany	1	3.8%	9.0%	0.7%	0.1%	Netherlands	1	0.0%	1.1%	0.3%	0.2%
	2	2.6%	8.1%	0.5%	0.0%		2	0.0%	0.3%	0.0%	0.0%
	3	0.0%	0.4%	0.1%	0.0%		3	0.0%	0.2%	0.4%	0.0%
	4	8.5%	11.2%	0.0%	0.0%		4	1.9%	3.6%	0.0%	0.0%
	5	8.2%	10.6%	0.0%	0.0%		5	0.8%	2.6%	0.0%	0.0%
	6	2.0%	4.0%	0.0%	0.0%		6	1.1%	2.2%	0.0%	0.0%
Denmark	1	2.2%	7.9%	3.3%	1.0%	Poland	1	0.0%	1.7%	0.8%	0.1%
	2	1.7%	7.0%	2.6%	0.8%		2	0.0%	0.0%	0.6%	0.1%
	3	0.1%	3.5%	1.9%	0.5%		3	0.0%	1.6%	0.6%	0.1%
	4	12.7%	14.6%	0.9%	0.0%		4	2.2%	3.6%	0.1%	0.0%
	5	12.8%	14.7%	0.7%	0.0%		5	0.0%	0.0%	0.0%	0.0%
	6	7.1%	11.0%	0.2%	0.0%		6	1.9%	2.9%	0.1%	0.0%
Estonia	1	2.9%	8.6%	0.9%	0.0%	Portugal	1	0.0%	0.0%	0.0%	0.0%
	2	2.9%	8.3%	0.3%	0.0%		2	0.0%	0.0%	0.0%	0.0%
	3	0.0%	0.5%	0.0%	0.0%		3	0.0%	0.0%	0.0%	0.0%
	4	8.1%	9.6%	0.0%	0.0%		4	0.0%	1.3%	0.0%	0.0%
	5	8.5%	8.9%	0.0%	0.0%		5	0.0%	0.1%	0.0%	0.0%
	6	1.5%	3.3%	0.0%	0.0%		6	0.0%	1.4%	0.0%	0.0%
Greece	1	0.0%	0.1%	0.0%	0.0%	Romania	1	1.8%	5.6%	0.1%	0.0%
	2	0.0%	0.1%	0.0%	0.0%		2	0.0%	0.5%	0.0%	0.0%
	3	0.0%	0.0%	0.0%	0.0%		3	0.0%	0.0%	0.0%	0.0%
	4	0.2%	1.3%	0.0%	0.0%		4	6.4%	8.8%	0.0%	0.0%
	5	0.2%	1.3%	0.0%	0.0%		5	1.0%	2.4%	0.0%	0.0%
	6	0.0%	0.5%	0.0%	0.0%		6	0.3%	3.3%	0.0%	0.0%

	Scenario	<0.3	<0.4	>3	>4		Scenario	<0.3	<0.4	>3	>4
Spain	1	1.3%	7.3%	0.3%	0.0%	Sweden	1	0.2%	2.4%	0.5%	0.0%
	2	0.6%	6.2%	0.1%	0.0%		2	0.0%	0.1%	0.1%	0.0%
	3	0.0%	0.0%	0.0%	0.0%		3	0.0%	1.2%	0.0%	0.0%
	4	8.5%	11.2%	0.0%	0.0%		4	3.7%	7.4%	0.0%	0.0%
	5	7.1%	11.9%	0.0%	0.0%		5	1.1%	2.7%	0.0%	0.0%
	6	0.0%	0.5%	0.0%	0.0%		6	1.5%	3.3%	0.0%	0.0%
Finland	1	11.0%	21.7%	2.9%	0.6%	Slovenia	1	0.0%	0.0%	0.0%	0.0%
	2	11.4%	21.7%	2.7%	0.6%		2	0.0%	0.0%	0.0%	0.0%
	3	4.3%	20.6%	2.5%	0.2%		3	0.0%	0.0%	0.0%	0.0%
	4	15.8%	21.1%	0.2%	0.0%		4	0.0%	0.4%	0.0%	0.0%
	5	16.0%	20.1%	0.2%	0.0%		5	0.0%	0.0%	0.0%	0.0%
	6	14.3%	19.8%	0.1%	0.0%		6	0.0%	0.3%	0.0%	0.0%
France	1	0.0%	0.5%	0.3%	0.0%	Slovakia	1	0.0%	1.9%	0.0%	0.0%
	2	0.0%	0.0%	0.2%	0.0%		2	0.0%	1.0%	0.0%	0.0%
	3	0.0%	0.0%	0.0%	0.0%		3	0.0%	0.2%	0.0%	0.0%
	4	1.6%	4.1%	0.0%	0.0%		4	3.7%	7.5%	0.0%	0.0%
	5	0.5%	1.7%	0.0%	0.0%		5	2.2%	7.6%	0.0%	0.0%
	6	0.0%	0.5%	0.0%	0.0%		6	1.9%	6.2%	0.0%	0.0%
Hungary	1	0.0%	0.6%	0.0%	0.0%	UK	1	12.6%	23.6%	4.8%	2.4%
	2	0.0%	0.2%	0.0%	0.0%		2	11.7%	22.6%	4.7%	2.1%
	3	0.0%	0.3%	0.0%	0.0%		3	0.9%	4.7%	0.5%	0.1%
	4	1.0%	2.5%	0.0%	0.0%		4	18.0%	23.4%	4.7%	0.3%
	5	0.8%	2.6%	0.0%	0.0%		5	18.1%	23.5%	4.6%	0.1%
	6	0.4%	1.0%	0.0%	0.0%		6	8.0%	10.2%	0.0%	0.0%

ad b). Impact on the sampling errors

Another criterion considered when choosing a scenario is the minimisation of the sampling errors. It is important to note that at this point we are not assessing the size of the sampling error itself, but its distribution depending on the selected scenario.

To carry out this analysis, we will focus on a subset of representative variables in 3rd EQLS, covering most of the objectives of the survey. This set was already proposed and discussed in the External Quality Assessment of 3rd EQLS[15] and cover a wide spectrum of the topics in the survey. Specifically, the variables to be considered are:

[15] www.eurofound.europa.eu/surveys/eqls/2011/documents/ef1337en.pdf

Q3, All things considered, how satisfied would you say you are with your life these days?
Please tell me on a scale of 1 to 1, where 1 means very dissatisfied and 10 means very satisfied. 98/99 DK/Ref

Q4, Could you please tell me on a scale of 1 to 10 how satisfied you are with each of the following items, where 1 means you are very dissatisfied and 10 means you are very satisfied? 98/99 DK/Ref

- b Your present job (HH2D=1/2)
- c Your present standard of living
- e Your family life

Q11. In general, do your working hours fit in with your family or social commitments outside work very well, quite well, not quite well or not at all well?
1 Very well, 2 Quite well, 3 Not quite well, 4 Not at all well

Q19. Do you have any of the following problems with your accommodation?

- b. Rot in windows, doors or floors
- c. Damp or leaks in walls or roof

1 Yes, 2 No

Q21. How frequently do you do each of the following?

- b. Use the Internet other than for work

1 Every day or almost every day, 2 At least once a week, 3 One to three times a month, 4 Less often, 5 Never

Q25. In all countries there sometimes exists tension between social groups.
In your opinion, how much tension is there between each of the following groups in this country?

- a. Poor and rich people
- c. Men and women
- g People with different sexual orientations

1 A lot of 2 Some 3 No Tension

Q27. Please look at the following statements about immigrants (i.e. people from abroad living in [COUNTRY]) and indicate where you would place your views on this scale. 98/99 Don't know/Refusal (DK/Ref)

- a. Immigrants are 1 not 10 well integrated in our society

Q41. Taking all things together on a scale of 1 to 1, how happy would you say you are? Here 1 means you are very unhappy and 10 means you are very happy. 98/99 DK/Ref

Q42. (Q43) In general, would you say your health is…?
1 Very good, 2 Good, 3 Fair, 4 Bad, 5 Very bad, 98/99 DK/Ref

Q45. Please indicate for each of the five statements which is closest to how you have been feeling over the last two weeks.

- a. I have felt cheerful and in good spirits

1 All of the time, 2 Most of the time, 3 More than half of the time, 4 Less than half of the time, 5 Some of the time, 6 At no time 98/99 DK/Ref

Q5, Please think about the area where you live now– I mean the immediate neighbourhood of your home. Do you have major, moderate or no problems with the following?

- a. Noise
- b. Air quality

1 Major problems, 2 Moderate problems, 3 No problems, 98/99 DK/Ref

Q53. (Q56) In general, how would you rate the quality of each of the following public services in COUNTRY? Please tell me on a scale of one to 1, where one means very poor quality and 10 means very high quality. 11/12 DK/Ref

- a. Health services
- b. Education system

Q57. Could you please evaluate the financial situation of your household? In comparison to most people in COUNTRY, would you say it is…?
1 Much worse, 2 Somewhat worse, 3 Neither worse nor better, 4 Somewhat better, 5 Much better 98/99 DK/Ref

Q58. A household may have different sources of income and more than one household member may contribute to it. Thinking of your household's total monthly income: is your household able to make ends meet….?
1 Very easily, 2 Easily, 3 Fairly easily, 4 With some difficulty, 5 With difficulty, 6 With great difficulty 98/99 DK/Ref

Q65. When you compare the financial situation of your household 12 months ago and now would you say it has become better, worse or remained the same?
1 Better, 2 The same, 3 Worse 97/98/99 NA/DK/Ref

CORE SOCIAL VARIABLES

Education
ISCED level 0-2
ISCED level 3-4
ISCED level 5-6

Activity status
Employed
Not Employed

Citizenship
Citizen of the country
Non-citizen of the country

As indicators of sampling error, coefficients of variation (as percentage) has been computed. The coefficient of variation (CV) is the ratio of the standard deviation to the mean, and as a result, it is a relative measure of precision. It can take only positive values and, as it does not depend on the unit of measurement, it can be calculated for all questions jointly. The lower the CV the smaller the relative sampling error. The formula used in this project is $CV = \frac{\sqrt{V(\hat{\hat{Y}})}}{\hat{\hat{Y}}} \cdot 100$ where $V\left(\hat{\hat{Y}}\right)$ is the sampling variance of the estimation and $\hat{\hat{Y}}$ is the estimation itself. The variances have been calculated by means of the usual unbiased estimates based on the sampling design that can be found in Lehtonen (2003)[16] or Lohr (1999)[17].

Box 3.1 Computation of sampling errors

Sampling errors have been computed by means of SPSS procedure CSTABULATE using the expressions below.[18]

Specifically, in the single-stage simple random sample (Malta) the variance of the total for a variable *y* is estimated as:

$$V(\hat{Y}) = \sum_{h=1}^{H} (1 - f_h) n_h S_h^2$$

where h is the stratum, H is the total number of strata, f_h the sampling fraction of stratum h, n_h the sample size in stratum h and S_h^2 the variance of stratum h.

For the rest of countries multi-stage sample without replacement at first stage was considered, in this case the variance is estimated as

$$V(\hat{Y}) = \sum_{h=1}^{H} (1 - f_h) n_h S_h^h + \sum_{h=1}^{H} \sum_{i=1}^{n_h} \pi_{hi} \sum_{k=1}^{K_{hi}} U_{hik}$$

where π_{hi} is the first stage inclusion probability for the primary sampling unit *i* in stratum *h*. K_{hi} is the number of second stage strata in the primary sampling unit *i* within the first stage stratum h. U_{hik} is variance contribution from the second stage stratum k from the primary sampling unit h_i. Its value depends on the second stage sampling method. Since sampling without replacement is considered, then

$$U_{hik=}(1 - f_{hi}) n_{hi} S_{hi}^2$$

where f_{hi} is the sampling fraction of stratum *k* from the primary sampling unit h_i, n_{hik} the sample size in stratum *k* from the primary sampling unit h_i and S_{hik}^2 the variance of stratum *k* from the primary sampling unit h_i.

16 Lehtonen, R. and Pahkinen, E. Practical methods for design and analysis of complex surveys, 2nd ed. John Wiley & Sons. 2003.
17 Lohr, S. Sampling: Design and Analysis. Duxbury Press. 1999.
18 IBM SPSS Statistics 20 Algorithms, IBM Corporation, 2011. pp. 238-239.

The information for the above 26 variables in all the countries is summarised in Figure 3.5[19]. For each scenario, this figure presents the distribution of the values of the coefficients of variation for all the answers to all the questions in all countries. The deciles of the coefficients of variation have been calculated and presented in Table 3.6. A comparison of the six scenarios shows that there is no relevant difference between the observed distributions of the coefficients of variation. Then, all six scenarios are equally appropriate as regards their impact on sampling errors. It should be noticed that the aim of this analysis is not an assessment of the values of the CV but a comparison of the impact of the weighting scenario on the distribution of such CVs.

Figure 3.5. Distribution of the coefficients of variation (EU27).

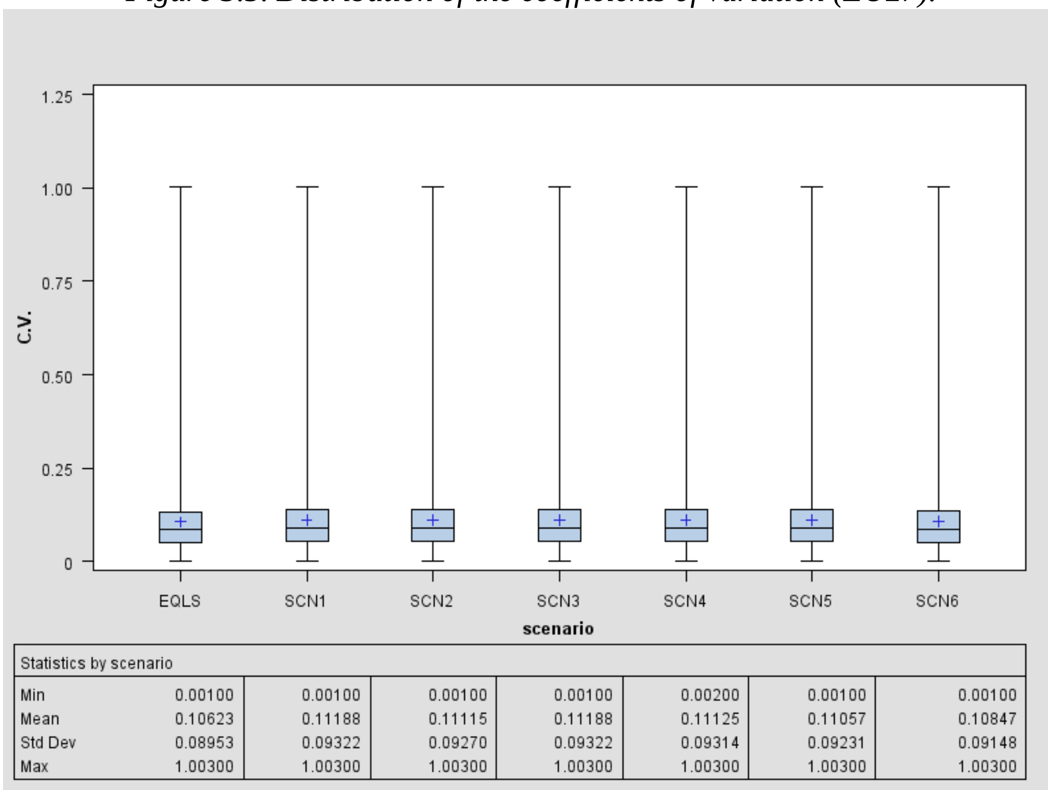


Table 3.6. Percentiles of the coefficients of variation by scenario (EU27).

	P05	P10	P25	P50	P75	P90	P95
EQLS	0.01950	0.02990	0.05092	0.08447	0.13205	0.20047	0.27590
Scenario1	0.01909	0.03095	0.05339	0.08959	0.14037	0.21616	0.28274
Scenario2	0.01901	0.03072	0.05316	0.08858	0.14013	0.21205	0.28341
Scenario3	0.01909	0.03095	0.05339	0.08959	0.14037	0.21616	0.28274
Scenario4	0.01907	0.03081	0.05317	0.08806	0.13924	0.21458	0.28439
Scenario5	0.01893	0.03057	0.05286	0.08785	0.13877	0.21211	0.28266
Scenario6	0.01852	0.03016	0.05178	0.08654	0.13531	0.20996	0.28109

[19] CV is a relative measure and high values (100%) could be expected in these cases where the denominator is low. For the variables considered in the analysis, a CV of 100% is only observed for the variable Citizenship in Hungary, where the estimation on non-nationals is 0.1% -0.2%, depending on the scenario.

As can be seen in tables 1 to 4 in Annex 1, same conclusion arise when a breakdown by country of the coefficients of variation is considered, the only exception is United Kingdom, where the CV for scenario 6 are slightly lower than the rest of scenarios (the mean of CV in UK is 0.079 for scenario 6, while for the rest of scenarios this mean lays between 0.093 and 0.097).

ad. c) Level of bias in those variables for which official data are available and are not included in the calibration procedure

In order check the level of bias generated by the different scenarios, the distribution of as sample of variables obtained with the different weights are compared with the distribution provided by official sources within the ESS, specifically EU-SILC and LFS. This analyse is carried out for the following variables[20]:

- *Education level*, with categories ISCED 0-2, ISCED 3-4 and ISCED 5-6 (comparison EQLS / LFS)
- *Nationality*, with categories 'citizen of the country' and 'non-citizen of the country' unemployed - employed (comparison EQLS / LFS)
- *Unemployment level*, with categories 'unemployed' and 'employed' (comparison EQLS / EU-SILC)
- *Q2. Are you mainly...?* with categories **'**Self-employed without employees', 'Self-employed with employees', 'Employed' and 'Other' (comparison EQLS / LFS)
- *Q19. Do you have any of the following problems with your accommodation? Rot in windows, doors or floors / Damp or leaks in walls or roof* with categories 'yes' and 'no' *Other'* (comparison EQLS / EU-SILC)
- *Q42 In general, would you say your health is …* with categories 'Very good', 'Good', 'Fair', 'Bad' and 'Very Bad' (comparison EQLS / EU-SILC)

To perform this comparison, a distance between the distributions of different variables depending on the scenario is defined in order to synthesize information and to facilitate analysis and decision based on this criterion. Whereas each of the percentage distribution of the variable as a vector, can be found in the literature a large number of distances that can be applied. Among them, the infinite distance, defined as

$$D_\infty(X,Y) = max_i\{|x_i - y_i|\}$$

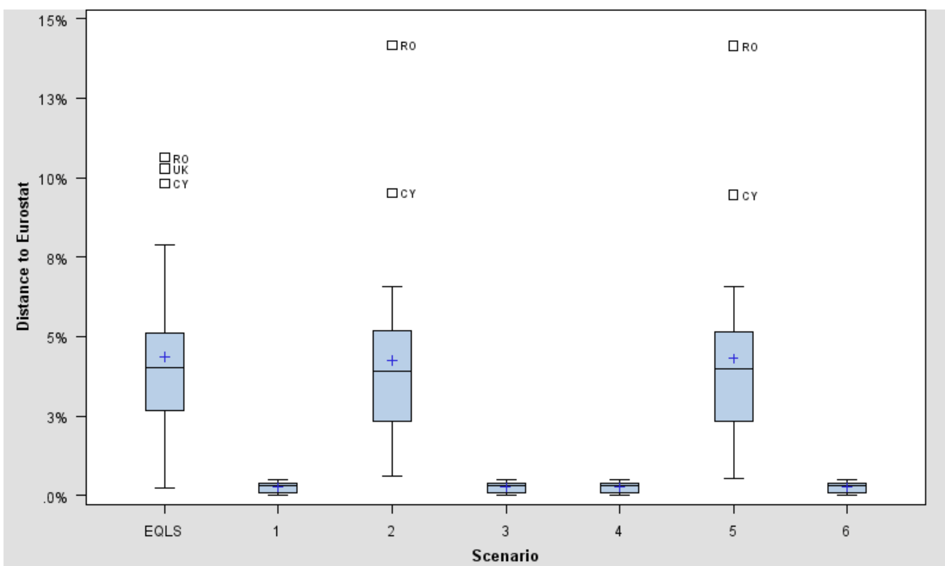
is chosen for simplicity of calculation and easy interpretation reasons. With this distance, we say that the nearest distribution to the reference distribution is the one with smaller maximum difference in their categories. While it is true that in this distance some information is lost, this is not too important, any distribution are in fact a vector of percentages and, consequently, their components are bounded. They can range between 0% and 100%. In any case, its easy interpretation outweighs the loss of information.

[20] The analysis has been carried out for all these variables that are in both (1) the subset of variables of 3rd EQLS specified above and (2) in LFS or EU-SILC. The goal of the analysis is to compare how the differences between EQLS and the other two surveys change in the different weighting scenarios. To achieve this goal, variables not included for calibration, such as nationality, are also considered in this section.

Figures 3.6 to 3.11 present the distribution of the distances between EQLS and the reference data for the six selected variables in all six weighting scenarios and the current weights used in 3rd EQLS. Despite of the obvious impact of the introduction of a variable in the calibration procedure, that eliminates the differences between the EQLS and reference distribution, on other relevant variation in the bias level is detected. This fact is supported by the similitude of the mean distances – as well as the persistence of the same outliers – among the different weighting scenarios.

The inclusion of a variable in a weighting scenario reduced the distances between the estimates of such a variable between 3rd EQLS and the reference survey (LFS or EU-SILC). This effect is observed in Figures 3.6 and 3.7. For instance, the variable activity status (Figure 3.6) is present in the calibration scenarios 1, 3, 4 and 6. In these scenarios, the calibration procedure has minimised the distance between the marginal distribution of activity status between EQLS and the reference survey and, as a consequence, the distribution of such a distance is more concentrated in the lower values, as shown by the box-plot graphs for scenarios 1,3,4 and 6 in figure 3.6.

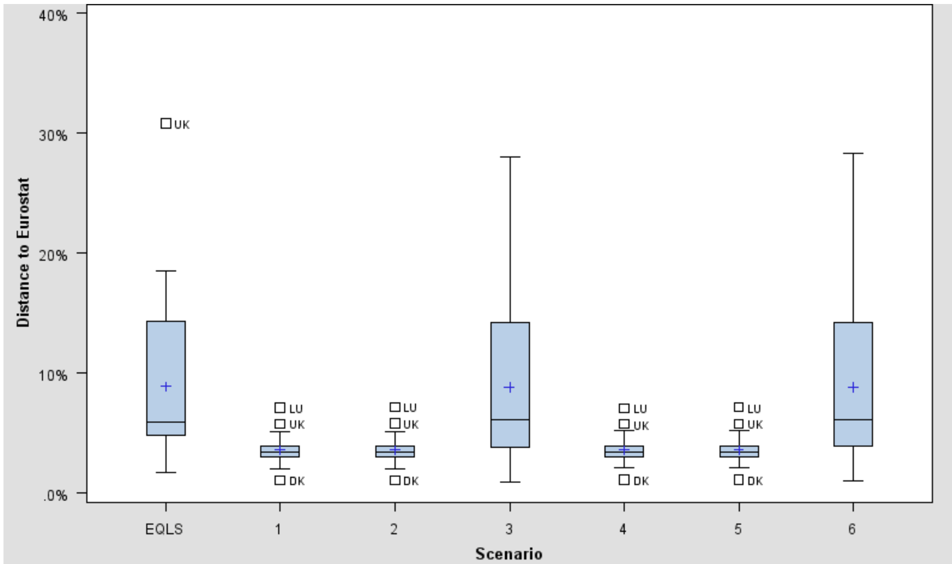
Figure 3.6. Distribution of the distance between the estimations of activity status in 3rd EQLS and EU-SILC 2012 (EU27).



Statistics by scenario							
Min	.22%	.03%	.60%	.03%	.03%	.53%	.03%
Mean	4.4%	.27%	4.3%	.27%	.27%	4.3%	.27%
Std Dev	2.7%	.15%	2.8%	.15%	.15%	2.7%	.15%
Max	11%	.49%	14%	.49%	.49%	14%	.49%

Activity

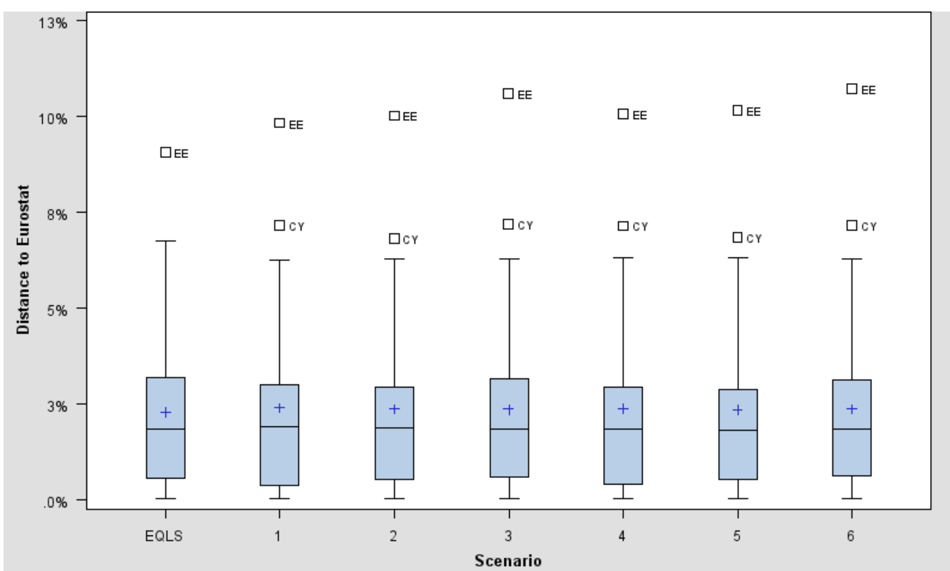
Figure 3.7. Distribution of the distance between the estimations of education level in 3rd EQLS and LFS 2012 (EU27).



Statistics by scenario							
Min	1.7%	1.1%	1.1%	.94%	1.2%	1.2%	.99%
Mean	8.9%	3.6%	3.6%	8.9%	3.6%	3.6%	8.8%
Std Dev	6.9%	1.2%	1.2%	6.7%	1.2%	1.2%	6.8%
Max	31%	7.2%	7.2%	28%	7.1%	7.2%	28%

Education

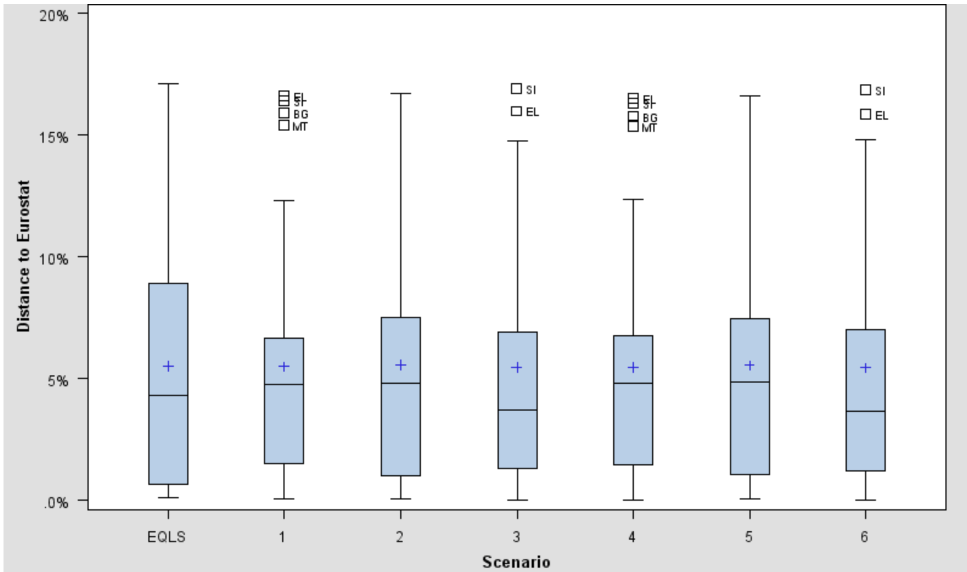
Figure 3.8. Distribution of the distance between the estimations of nationality in 3rd EQLS and EU-SILC 2012 (EU27).



Statistics by scenario							
Min	.05%	.05%	.04%	.05%	.05%	.04%	.05%
Mean	2.3%	2.4%	2.4%	2.4%	2.4%	2.4%	2.4%
Std Dev	2.2%	2.4%	2.4%	2.4%	2.4%	2.4%	2.5%
Max	9.1%	9.8%	10%	11%	10%	10%	11%

Nationality

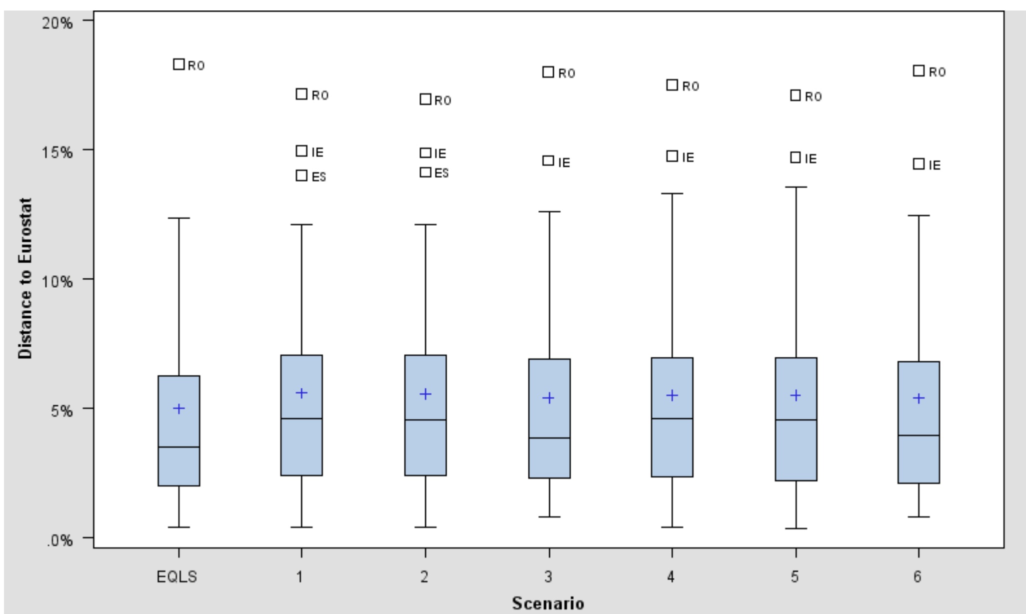
Figure 3.9. Distribution of the distance between the estimations of Q19 in 3rd EQLS and EU-SILC 2012 (EU27).



Statistics by scenario	EQLS	1	2	3	4	5	6
Min	.10%	.07%	.09%	.02%	.02%	.07%	.00%
Mean	5.5%	5.5%	5.6%	5.5%	5.5%	5.6%	5.5%
Std Dev	5.2%	5.4%	5.4%	5.2%	5.4%	5.4%	5.2%
Max	17%	17%	17%	17%	17%	17%	17%

Q19bc

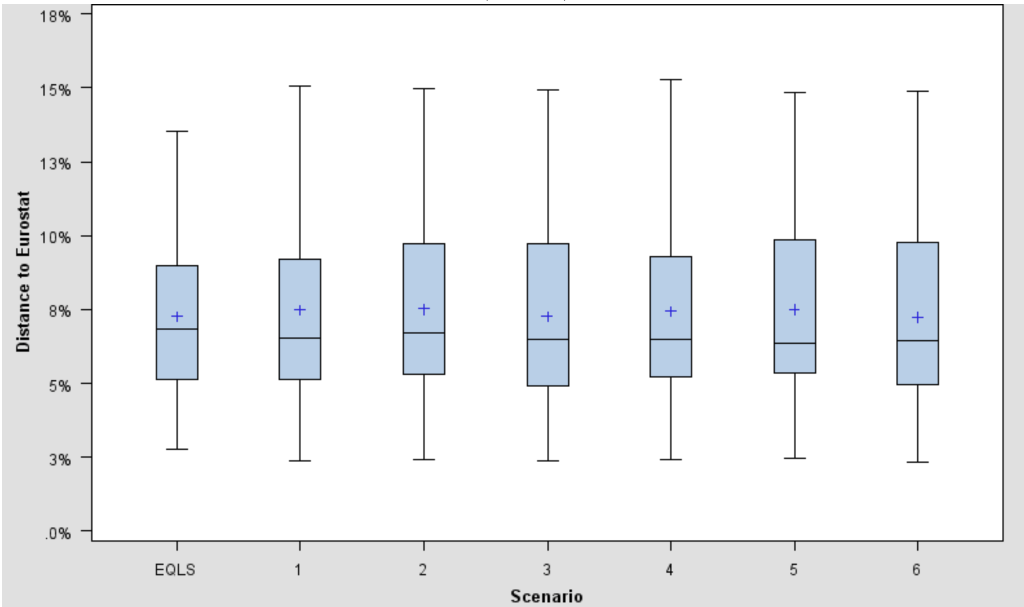
Figure 3.10. Distribution of the distance between the estimations of Q2 in 3rd EQLS and EU-SILC 2012 (EU27).



Statistics by scenario	EQLS	1	2	3	4	5	6
Min	.41%	.43%	.43%	.80%	.40%	.39%	.83%
Mean	5.0%	5.6%	5.6%	5.4%	5.5%	5.5%	5.4%
Std Dev	4.3%	4.5%	4.5%	4.4%	4.4%	4.4%	4.4%
Max	18%	17%	17%	18%	18%	17%	18%

Q2

Figure 3.11. Distribution of the distance between the estimations of Q42 in 3rd EQLS and EU-SILC 2012 (EU27).



Statistics by scenario							
Min	2.8%	2.4%	2.4%	2.4%	2.4%	2.5%	2.4%
Mean	7.3%	7.5%	7.5%	7.3%	7.4%	7.5%	7.2%
Std Dev	3.1%	3.3%	3.3%	3.2%	3.3%	3.3%	3.2%
Max	14%	15%	15%	15%	15%	15%	15%

Q42

Despite figures 3.6 to 3.11 show very similar distributions of the distance among scenarios it can be seen that scenarios 1 and 4 (including both education level and activity status) are closer to the reference distributions for all variables excepting Q2. Since the inclusion of two additional variables does not increase the distortion of the design weight (see subsection 3.2.1), these scenarios could be considered as more appropriate for the calibration strategy of EQLS.

ad. d) Level of variation of the estimations obtained with the current calibration methodology of EQLS

Table 3.6 presents the differences among the distribution of the subset of questions in 3rd EQLS that cannot be compared with an external reference in the ESS, since these variables are not included in EU-SILC or LFS. Specifically, the table shows the difference of the estimates in each of the six weighting scenarios with respect to the current calibration of 3rd EQLS.

The main conclusion from this table is that the changes in the distribution are very small (always in the second decimal digit). This robustness in the estimations is also found when the variables are compared at a country level[21]. This result is quite positive, since if the distances between the original EQLS estimates and the correspondent estimates

[21] Since all the figures from a comparison country by country do not provide with any additional information, they are not included in the report.

produced by each of the scenarios were large, it would not be possible to decide which of the estimations is the best one due to the lack of reference data[22]. However, if the differences are not large, as it is the case, no relevant problem in for the use of any of the scenarios arises.

3.2.5. Main conclusion

The main conclusion of this section is that the impact of the application of the different weighting scenarios does not generate important changes in the estimated distribution of the analysed variables. This robustness is a very positive feature of 3rd EQLS and, specifically, of its current sampling and calibration methodologies. Considering all selection criteria altogether (shown in Figure 3.12), scenario 4 seems to be the most appropriate for the calibration. It is the scenario with lower distances to external information along with scenario 1, but scenario 1 has much larger distortion of the calibrated weights (a problem that gets worse for some countries as UK, Denmark or Poland).

Figure 3.12. Summary of the results of the comparison criteria.

	Scenario					
	1	2	3	4	5	6
A. Level of distortion of the design weights	▼	▼	–	–	–	▲
B. Impact on the sampling errors	–	–	–	–	–	–
C. Level of bias in those variables for which official data are available	▲	▼	–	▲	▼	–
D. Level of variation of the estimations	–	–	–	–	–	–

Scenario 4 has the disadvantage of leading to some negative weights. However, only 1,5% of the cases have a negative weight. Moreover, in these countries with the largest number of negative weights – such as UK - scenarios using IPF generate a larger distortion of the design weights than scenario 4 and should no be recommended for the weighting strategy of EQLS. The issue of the negative weights in Scenario 4 will be discussed in the next section, where a solution is proposed.

Table 3.7. Differences between the estimation for each calibration scenario and the current calibration of 3rd EQLS (EU27 - percentage points).

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Q11	Very well	0.02	0.02	0.02	0.02	0.02	0.02
	Fairly well	0.01	0.01	0.01	0.01	0.01	0.01
	Not very well	0.03	0.03	0.03	0.03	0.03	0.02
	Not at all well	0.05	0.05	0.05	0.05	0.05	0.04
Q19b	Yes	0.03	0.03	0.03	0.03	0.03	0.03
	No	0.00	0.00	0.00	0.00	0.00	0.00
Q19c	Yes	0.02	0.02	0.02	0.02	0.02	0.02
	No	0.00	0.00	0.00	0.00	0.00	0.00

[22] The lack of reference information in the ESS establishes the difference between this criterion and that discussed at subsection 3.2.3.

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Q21b	Every day or almost	0.01	0.01	0.01	0.01	0.01	0.01
	At least once a week	0.02	0.02	0.02	0.02	0.02	0.02
	One to three times a month	0.05	0.05	0.05	0.05	0.05	0.04
	Less often	0.04	0.04	0.04	0.04	0.04	0.04
	Never	0.01	0.01	0.01	0.01	0.01	0.01
Q25a	A lot of tension	0.01	0.01	0.01	0.01	0.01	0.01
	Some tension	0.01	0.01	0.01	0.01	0.01	0.01
	No tension	0.03	0.03	0.03	0.03	0.03	0.03
Q25g	A lot of tension	0.02	0.02	0.02	0.02	0.02	0.02
	Some tension	0.01	0.01	0.01	0.01	0.01	0.01
	No tension	0.02	0.02	0.02	0.02	0.02	0.02
Q27a	1- Immigrants are not integrated in our society	0.04	0.03	0.04	0.04	0.03	0.03
	2	0.03	0.03	0.03	0.04	0.03	0.03
	3	0.03	0.03	0.03	0.02	0.02	0.02
	4	0.02	0.02	0.02	0.02	0.02	0.02
	5	0.02	0.01	0.02	0.01	0.01	0.01
	6	0.02	0.02	0.02	0.02	0.02	0.02
	7	0.02	0.02	0.02	0.02	0.02	0.02
	8	0.03	0.03	0.03	0.03	0.03	0.03
	9	0.06	0.06	0.06	0.05	0.05	0.05
	10-Immigrants are well integrated in our society	0.05	0.05	0.05	0.05	0.05	0.05
Q42	Very good	0.02	0.02	0.02	0.02	0.02	0.02
	Good	0.01	0.01	0.01	0.01	0.01	0.01
	Fair	0.01	0.01	0.01	0.01	0.01	0.01
	Bad	0.03	0.03	0.03	0.03	0.03	0.03
	Very bad	0.05	0.05	0.05	0.05	0.05	0.05
Q45a	All of the time	0.03	0.03	0.03	0.03	0.03	0.02
	Most of the time	0.01	0.01	0.01	0.01	0.01	0.01
	More than half of the time	0.02	0.02	0.02	0.02	0.02	0.01
	Less than half of the time	0.02	0.02	0.02	0.02	0.02	0.02
	Some of the time	0.03	0.03	0.03	0.03	0.03	0.03
	At no time	0.05	0.05	0.05	0.05	0.05	0.05
Q50a	Major problems	0.04	0.04	0.04	0.04	0.04	0.04
	Moderate problems	0.02	0.02	0.02	0.02	0.02	0.02
	No problems	0.01	0.01	0.01	0.01	0.01	0.01
Q50b	Major problems	0.05	0.05	0.05	0.05	0.05	0.05
	Moderate problems	0.02	0.02	0.02	0.02	0.02	0.02
	No problems	0.01	0.01	0.01	0.01	0.01	0.01
Q57	Much worse	0.04	0.04	0.04	0.04	0.04	0.04
	Somewhat worse	0.02	0.02	0.02	0.02	0.02	0.02
	Neither worse nor better	0.01	0.01	0.01	0.01	0.01	0.01
	Somewhat better	0.02	0.02	0.02	0.02	0.02	0.02
	Much better	0.04	0.04	0.04	0.04	0.04	0.03

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Q58	Very easily	0.03	0.03	0.03	0.03	0.03	0.03
	Easily	0.02	0.02	0.02	0.02	0.02	0.02
	Fairly easily	0.01	0.01	0.01	0.01	0.01	0.01
	With some difficulty	0.01	0.01	0.01	0.01	0.01	0.01
	With difficulty	0.03	0.03	0.03	0.03	0.03	0.02
	With great difficulty	0.03	0.03	0.03	0.03	0.03	0.03
Q65	Better	0.03	0.02	0.03	0.02	0.02	0.02
	The same	0.01	0.01	0.01	0.01	0.01	0.01
	Worse	0.01	0.01	0.01	0.01	0.01	0.01

Chapter 4: Strategies for treatment of extreme weights

Once a calibration scenario has been selected, two additional issues related to extreme values of the weights should be considered:

- Since scenario 4 applies the GREG estimation method, some negative weights can appear. The interpretation and application of negative weights could be difficult and may generate potential errors by unskilled users. Moreover, if the number of negative weights were very high, some subpopulation estimate could be higher than that of the total population. For these reasons, negatives weights need to be corrected in this chapter.
- Even being positive, the concern of applying too large weights is that some features specific of a small subsample can be actually extrapolated to a large subpopulation. To avoid this issue, extreme weights should be treated.

4.1. Treatment of negative weights

Considering that a negative weight is hard to explain and interpret, positivity of weights is considered as a must. A first ad-hoc solution to the problem of existence of negative weights is to set a fixed small and positive value for all of them. However, this type of trimming practices is arbitrary and may generate a misalignment between the estimates with the new weights and actual marginal distribution, invalidating the whole calibration procedure.

As discussed in Subsection 2.3, any calibration problem can be solved, in mathematical terms, as a minimization problem (with objective function as the distance between original weights and calibrated weights) with restrictions. In this situation, a better solution is to take profit that the general framework of calibration is equivalent to a minimization problem (distance between design and calibrated weights) subject to certain restrictions (estimates using calibrated weights have to match some given totals) only adding as additional restriction that the calibrated weights have to be positive or (equivalently) the ratio between final and initial weights greater than zero. That is, to ensure that the new weights are positive it is only necessary to repeat exactly the same calibration process which has already been carried out for GREG but including the additional restriction that the ratio between the calibrated weights and design is larger than a strictly positive lower bound[23]. This is the approach followed in our analysis, where a lower bound equal to 0.1 has been defined because only a small proportion of weights (between 0.2% and 0.7% depending on scenario) is included in the interval [0, 0.1] . Besides that the standard restriction specifying that total estimated subpopulations match the actual marginal distributions of the calibration variables is also included. This procedure creates a new set of weights, wb_4, which will be used in the analysis hereafter.

[23] Zero cannot be considered an appropriated bound because it could result in null weights to some observations. The additional constrain was not introduced from the very beginning in order to compare the six scenarios in the same conditions. Once a scenario was selected in this comparison stage, the additional constrain is introduced as part of the strategy to manage extreme weights.

To assess the impact of the new restrictions in the optimisation problem, the ratio between the new weights with the non-negativity constrain (*wb_4*,) and the original weights for scenario 4 is considered. Table 4.1 presents the distribution of this ratio.

$$Ratio = wb_4 / w_{GREG\_4}$$

Table 4.1. Distribution of the ratio of wb_4 over the original weight obtained in scenario 4 (Percentiles).

P05	P10	P25	P50	P75	P90	P95
0.98	0.99	1.00	1.00	1.00	1.03	1.11

By looking at the distribution of the calculated ratio, we can see that the inclusion of the bound has modified the weights, but due to the small prevalence of negative weights in scenario 4 (1.5 %), more than 90% of observed ratios take values very close to 1 and wb_4 can be safely applied instead of the original weights.

4.2. Treatment of positive extreme weights

The usual goal of trimming is to reduce the impact of extreme weights in the variance of the estimates, while trying to introduce the least possible bias, due to the distortion of trimmed weights[24]. That is to say, the idea is to reduce the variance without increasing the bias too much, reaching a balance between these two opposing effects.

Traditionally, most applications of trimming finally arrive to the establishment of one or more arbitrarily fixed bounds to the weights in order to try to meet those conditions[25]. In this report, the consultant suggest the application of an alternative method – namely a shrinkage method, to achieve the same goal but trying to eliminate the element of arbitrariness in the choice of what are the extreme weights that are implicit in commonly used trimming strategies.

The bias variance balance can be conveniently characterized by the mean square error. Therefore, the problem can be dealt with much more precise terms as that of obtaining a new set of weights, from the calibrated weights that minimize the mean square error of the estimates. That is precisely the reason that led to choose the methodology of shrinkage[26]. An approach that takes more profit from the information contained in the sample and which basically consists of obtaining a new set of weights through a convex combination of weights before calibration and the calibrated weights (in our case w_0 and wb_4, respectively). Analytically:

$$final\ weight = \lambda \cdot wb_4 + (1 - \lambda) \cdot w_0$$

where λ is a value between 0 and 1 calculated from the sample that minimize the mean squared error of the estimate.

[24] Weighting for Unequal Pi. Kish, L. (1992). Journal of Official Statistics, Vol.8, No.2,
[25] Potter, F.J. (1988) Survey Procedures to Control Extreme Sampling Weights, Porceedings of the Section on Survey Research Methods, American Statistical Association
[26] Shrinkage weights for unequal probability samples (1991). Cohen, T. and Spencer, B.D. Proceedings of the sections on Survey Research Methods, American Statitistical Association

Specifically, according to the methodology of the article, the obtained λ satisfies the following properties:

- it has less bias than the estimator before calibration
- it has less variance than the calibration estimator
- it has less mean squared error than either of the two estimators

λ can be estimated explicitly from sample estimates of variances, covariances and biases as detailed in Cohen and Spencer (1991)[27]. The formula used for λ calculation is the following:

$$\lambda = \begin{cases} \min\left(\frac{a_{b4}}{a_0 + a_{b4}}; 1\right) & , a_{b4} \geq 0, a_{b4} + a_0 \geq 0 \\ 0 & , \text{otherwise} \end{cases}$$

being,

X_{b4} the estimation using the weight w_{b4}

X_0 the estimation using the weight w_0

$$a_{b4} = \widehat{MSE}(X_{b4}) - \widehat{Cov}(X_{b4}; X_0)$$

$$a_0 = \hat{V}(X_0) - \widehat{Cov}(X_{b4}; X_0)$$

$$\widehat{MSE}(X_{b4}) = \hat{V}(X_{b4}) + \hat{B}(X_{b4})^2$$

$$\hat{B}(X_{b4})^2 = (X_{b4} - X_0)^2 - \hat{V}(X_0) - \hat{V}(X_{b4}) + 2\widehat{Cov}(X_{b4}; X_0)$$

The application of shrinkage method could generate a different λ for each variable and country, which would make it impractical. An operational solution is to consider estimates for the total across all countries, analysing the set of variables defined in Chapter 3. This generates 57 different values of λ, one for each category of the variables, which are listed in Table 4.2, where the variable and the value of λ are shown.

[27] Shrinkage weights for unequal probability samples (1991). Cohen, T. and Spencer, B.D. Proceedings of the sections on Survey Research Methods, American Statitistical Association

Table 4.2. Optimal value of λ for each variable and category.

Variable and Category	λ	Variable and Category	λ	Variable and Category	λ
q11	0.999	q21b_1	0.974	q36b_3	0.964
q27a	0.000	q21b_2	0.820	q36b_4	0.951
q30	0.000	q21b_3	0.954	q36b_5	0.927
q40b	0.999	q21b_4	0.931	q45a_1	0.817
q40e	0.974	q21b_5	0.989	q45a_2	0.148
q41	0.589	q25a_1	0.814	q45a_3	0.496
q42	0.998	q25a_2	0.000	q45a_4	0.960
q53a	0.314	q25a_3	0.099	q45a_5	0.847
q53b	0.000	q25b_1	0.893	q45a_6	0.255
q57	0.000	q25b_2	0.000	q50a_1	0.895
q58	0.988	q25b_3	0.229	q50a_2	0.000
q2_1	0.956	q25c_1	0.864	q50a_3	0.000
q2_2	0.928	q25c_2	0.000	q50b_1	0.000
q2_3	0.995	q25c_3	0.000	q50b_2	0.885
q2_4	0.923	q25g_1	0.000	q50b_3	0.000
q19b_1	0.000	q25g_2	0.580	q65_1	0.975
q19b_2	0.977	q25g_3	0.393	q65_2	0.911
q19c_1	0.000	q36b_1	0.997	q65_3	0.973
q19c_2	0.923	q36b_2	0.970		

As it can be seen in the above table, there are clearly two distinct groups of values of λ depending on whether the optimal is or not greater than 0. The zero value cases are probably due to the poor relationship between the calibration variables and the corresponding categories. On the other hand, the values in the other group tend to cluster towards a value close to 1. In this case, as a compromise, it was decided to select as an overall λ the median of all the values, which can be considered a robust approximation to a hypothetical general optimum. In particular $\lambda = 0.86$.

The main effect of shrinkage is a 'concentration' of the distribution the weights, leading towards a reduction of the prevalence of extreme weights. Such an effect is confirmed by Figure 4.1 and Table 4.3, which present the ratio between wb_4 (before and after shrinkage) and the design weight. Notice that all percentiles of weight after shrinkage are closer to 1 than before the shrinkage transformation. Then, the application of the weights with shrinkage leads to estimators with a lower mean squared error.

Figure 4.1. Distribution of the ratio of the weights before shrinkage to the design weights and after shrinkage to the design weights (EU27)

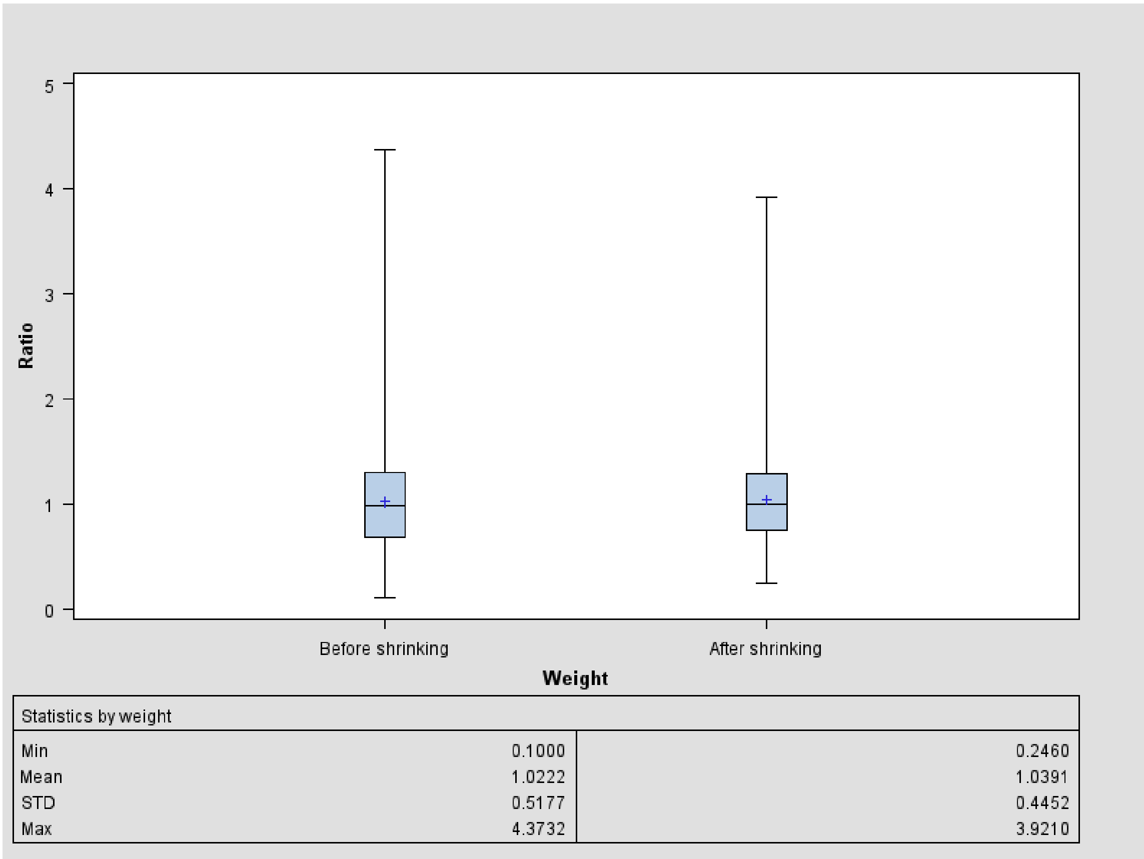


Table 4.3. Distribution of the ratio of the weights before shrinkage to the design weights and after shrinkage to the design weights (EU27-Percentiles)

	P05	P10	P25	P50	P75	P90	P95
Ratio after shrinkage	0.33	0.51	0.72	0.98	1.26	1.57	1.81
Ratio before shrinkage	0.22	0.43	0.68	0.98	1.30	1.66	1.94

At a country level analysis, Figure 4.2 shows that the shrinkage process do not generate any relevant distortion of the weight distribution in any of the EU27 Member States. Shrinkage just lightly increases the relative frequency of central weights with a reduction of the extreme ones optimizing the trade-off between bias and variance to minimize the squared error of the estimations.

Figure 4.2. Distribution of weights before and after shrinkage by country

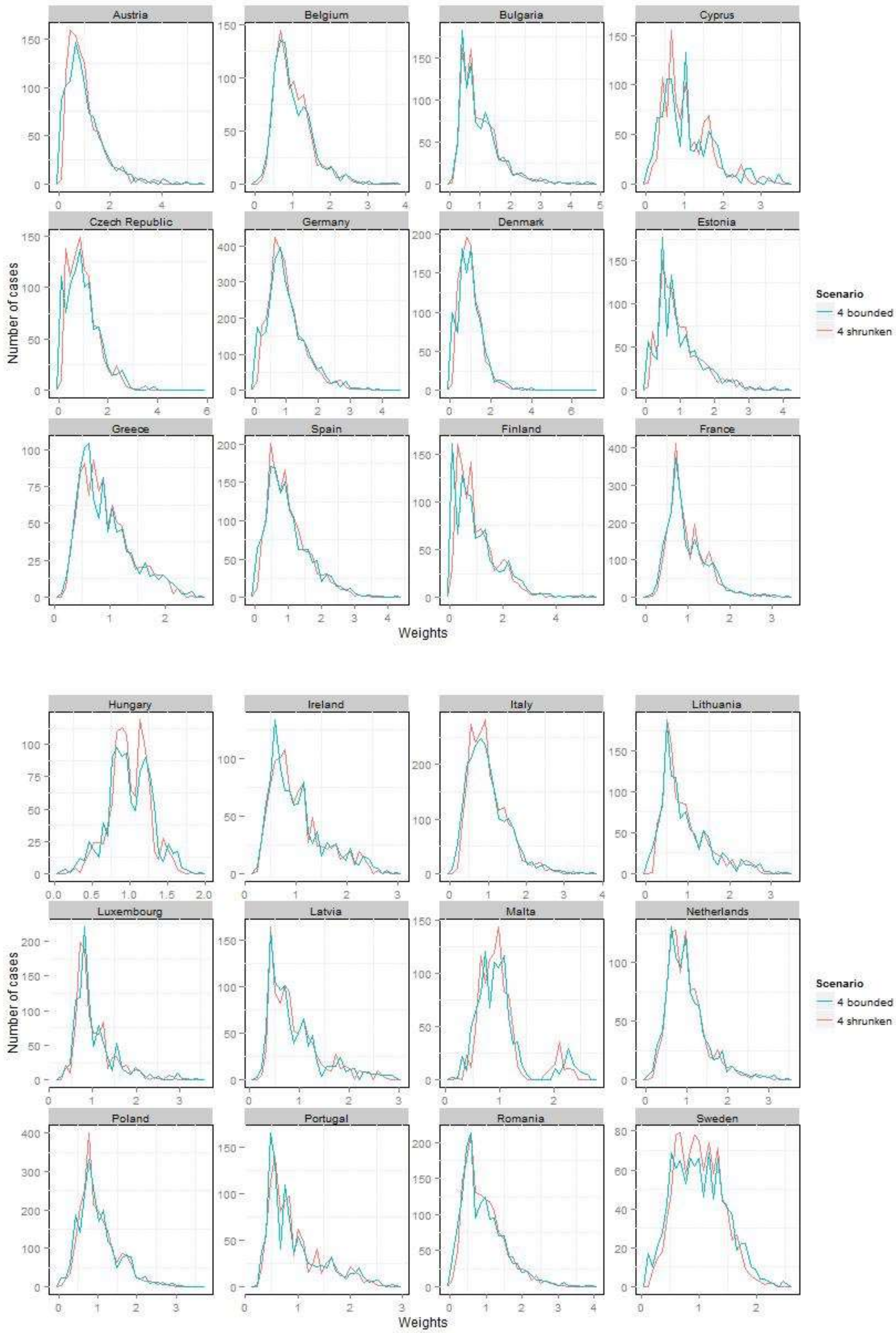
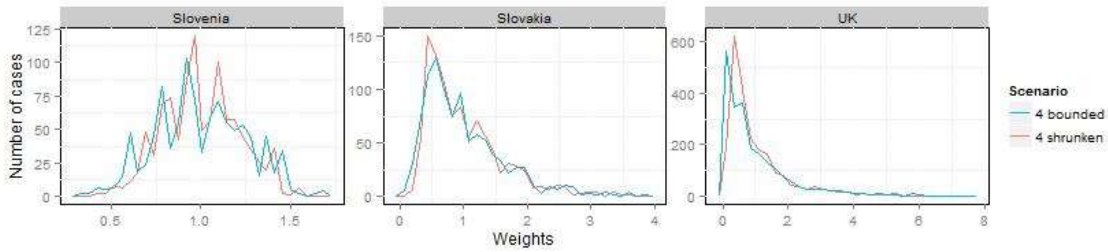


Figure 4.2. Distribution of weights before and after shrinkage by country (continuation)



After the treatment of the extreme weights, negative and extreme large (>3) do not represent an issue (Table 4.4.). Extremely small weights (<0.3) are applied to 4.1% of the cases. However, 30% of these cases with extremely low weights correspond to UK and almost half of them (46.5%) are concentrated in Czech Republic, Poland and UK[28]. Since small weight cannot artificially amplify outliers at the population level (as extremely high weights can), no additional treatment of extreme weight (such as trimming) is recommended.

Table 4.4. Extreme weights for scenario 4, before and after their treatment (EU27)

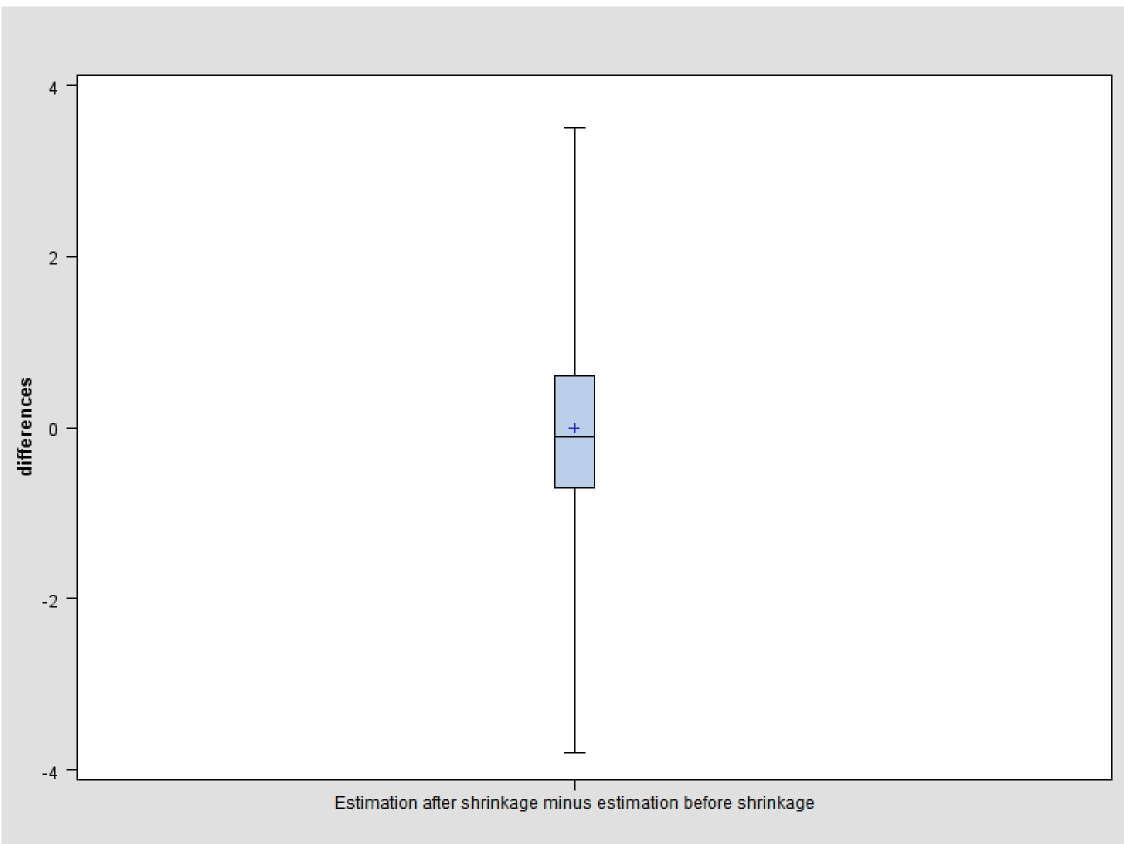
Scenario		Negative	<0.3	<0.4	>3	>4
Total	**4 (bounded and shrunk)**	**0.0%**	**4.1%**	**6.1%**	**0.2%**	**0.0%**
	4 (initial)	1.4%	5.5%	8.3%	0.4%	0.0%
Austria	4 (bounded and shrunk)	0.0%	7.8%	10.9%	0.0%	0.0%
	4 (initial)	2.0%	10.8%	17.7%	0.1%	0.0%
Belgium	4 (bounded and shrunk)	0.0%	0.4%	1.0%	0.0%	0.0%
	4 (initial)	0.0%	1.5%	2.9%	0.0%	0.0%
Bulgaria	4 (bounded and shrunk)	0.0%	0.4%	1.2%	0.0%	0.0%
	4 (initial)	0.0%	1.4%	3.9%	0.1%	0.0%
Cyprus	4 (bounded and shrunk)	0.0%	4.0%	5.4%	0.0%	0.0%
	4 (initial)	0.9%	5.6%	10.5%	0.0%	0.0%
Czech Republic	4 (bounded and shrunk)	0.0%	9.3%	12.6%	0.4%	0.0%
	4 (initial)	3.9%	11.5%	15.3%	1.1%	0.0%
Germany	4 (bounded and shrunk)	0.0%	6.9%	9.4%	0.0%	0.0%
	4 (initial)	2.8%	8.5%	11.2%	0.0%	0.0%
Denmark	4 (bounded and shrunk)	0.0%	8.9%	13.1%	0.5%	0.0%
	4 (initial)	2.4%	12.7%	14.6%	0.9%	0.0%
Estonia	4 (bounded and shrunk)	0.0%	7.4%	9.1%	0.0%	0.0%
	4 (initial)	2.3%	8.1%	9.6%	0.0%	0.0%

[28] The equivalences between the national classifications for education in these countries and ISCED levels could be revised for future waves of EQLS.

Scenario		Negative	<0.3	<0.4	>3	>4
Greece	4 (bounded and shrunk)	0.0%	0.0%	0.2%	0.0%	0.0%
	4 (initial)	0.0%	0.2%	1.3%	0.0%	0.0%
Spain	4 (bounded and shrunk)	0.0%	4.6%	8.2%	0.0%	0.0%
	4 (initial)	1.4%	8.5%	11.2%	0.0%	0.0%
Finland	4 (bounded and shrunk)	0.0%	14.8%	19.4%	0.1%	0.0%
	4 (initial)	6.5%	15.8%	21.1%	0.2%	0.0%
France	4 (bounded and shrunk)	0.0%	0.4%	1.3%	0.0%	0.0%
	4 (initial)	0.0%	1.6%	4.1%	0.0%	0.0%
Hungary	4 (bounded and shrunk)	0.0%	0.5%	0.7%	0.0%	0.0%
	4 (initial)	0.0%	1.0%	2.5%	0.0%	0.0%
Ireland	4 (bounded and shrunk)	0.0%	0.0%	0.0%	0.0%	0.0%
	4 (initial)	0.0%	0.0%	0.1%	0.0%	0.0%
Italy	4 (bounded and shrunk)	0.0%	1.5%	3.3%	0.0%	0.0%
	4 (initial)	0.0%	3.8%	7.8%	0.3%	0.0%
Lithuania	4 (bounded and shrunk)	0.0%	4.4%	6.0%	0.0%	0.0%
	4 (initial)	0.4%	6.1%	10.5%	0.0%	0.0%
Luxembourg	4 (bounded and shrunk)	0.0%	0.0%	0.0%	0.0%	0.0%
	4 (initial)	0.0%	0.0%	0.0%	0.0%	0.0%
Latvia	4 (bounded and shrunk)	0.0%	0.5%	3.0%	0.0%	0.0%
	4 (initial)	0.0%	3.2%	5.9%	0.0%	0.0%
Malta	4 (bounded and shrunk)	0.0%	0.3%	0.7%	0.0%	0.0%
	4 (initial)	0.0%	1.5%	3.2%	0.0%	0.0%
Netherlands	4 (bounded and shrunk)	0.0%	0.6%	1.6%	0.0%	0.0%
	4 (initial)	0.1%	1.9%	3.6%	0.0%	0.0%
Poland	4 (bounded and shrunk)	0.0%	1.2%	2.0%	0.0%	0.0%
	4 (initial)	0.0%	2.2%	3.6%	0.1%	0.0%
Portugal	4 (bounded and shrunk)	0.0%	0.0%	0.0%	0.0%	0.0%
	4 (initial)	0.0%	0.0%	1.3%	0.0%	0.0%
Romania	4 (bounded and shrunk)	0.0%	3.5%	6.4%	0.0%	0.0%
	4 (initial)	1.9%	6.4%	8.8%	0.0%	0.0%
Sweden	4 (bounded and shrunk)	0.0%	1.8%	3.5%	0.0%	0.0%
	4 (initial)	0.3%	3.7%	7.4%	0.0%	0.0%
Slovenia	4 (bounded and shrunk)	0.0%	0.0%	0.0%	0.0%	0.0%
	4 (initial)	0.0%	0.0%	0.4%	0.0%	0.0%
Slovakia	4 (bounded and shrunk)	0.0%	0.3%	3.4%	0.0%	0.0%
	4 (initial)	0.0%	3.7%	7.5%	0.0%	0.0%
UK	4 (bounded and shrunk)	0.0%	19.5%	24.6%	3.0%	0.0%
	4 (initial)	8.0%	18.0%	23.4%	4.7%	0.3%

The assessment of the treatment of extreme weights needs to be completed with an analysis of the impact of such a treatment in the value of the estimations. Figure 4.3 exhibits the differences between the estimations before and after shrinkage, where each case is defined as a combination of question by category and by country. The interquartile range, where lies 50% of the sample, is 1.3 points. It is important to remark that the distribution shown in Figure 4.3 is symmetric and these differences are not high biased in either way. The presence of a bias in this distribution could be interpreted as a signal of the lack of reliability of the shrinkage procedure, that would not only reduce the extreme weights (as required) but also distort the design weights re-assigning probability mass from the cases in some strata to those in a different stratum.

Figure 4.3. Distribution of the differences between the estimates after and before shrinkage (EU27-Percentage points).



Finally, Table 4.5 presents the estimations obtained by applying the current weights in 3rd EQLS, the weights obtained in scenario 4 with the additional constrain of non-negativity in the optimisation problem (wb_4) and the latter weight after shrinkage. The table shows no large differences between estimators before and after shrinkage in all questions but Q42. Even for this variable, the difference is lower than 3%. In other words, the application of shrinkage weights does introduce relevant changes in the estimations and reduces the presence of extreme positive weights.

Table 4.5. Example of estimations using the current 3rd EQLS weights, wb_4 and wb_4 after shrinkage (EU27).

		Weight after shrinkage	Weight before shrinkage	current weight of 3rd EQLS
Q2. Are you mainly...?	Self-employed without employees	11.3%	11.7%	12.5%
	Self-employed with employees	4.1%	4.4%	4.1%
	Employed	82.1%	80.7%	80.7%
	Other	2.5%	3.1%	2.7%
Q19b. Rot in windows, doors or floors / Do you have any of the following problems with your accommodation?	Yes	9.7%	8.3%	10.6%
	No	90.3%	91.7%	89.4%
Q19c. Damp or leaks in walls or roof / Do you have any of the following problems with your accommodation?	Yes	13.1%	11.5%	13.9%
	No	86.9%	88.5%	86.1%
Q42 In general, would you say your health is …	Very good	21.4%	23.1%	22.1%
	Good	38.9%	42.2%	42.3%
	Fair	29.0%	26.2%	26.4%
	Bad	8.4%	6.6%	7.0%
	Very bad	2.3%	1.9%	2.2%
Activity status	Unemployed	11.6%	11.1%	15.1%
	Employed	88.4%	88.9%	84.9%
Education level	ISCED level 0-2	32.5%	33.0%	36.8%
	ISCED level 3-4	44.3%	44.1%	41.9%
	ISCED level 5-6	23.2%	22.9%	21.3%
Nationality	Non-citizen of the country	4.7%	4.3%	4.2%
	Citizen of the country	95.3%	95.7%	95.8%

4.3. Conclusion

The main limitation for the application of the GREG calibration method – the presence of a small (1.5%) proportion of cases with negative weights - can be easily solved with the introduction of additional non-negativity constrains in the underlying optimisation problem. The weights obtained under these constrains are always positive and do not generate any relevant distortion.

The application of ad hoc trimming procedures that may be arbitrary and not supported by any theory is not recommended by the literature[29]. As alternative, the shrinkage methodology applied in this subsection is based on an optimal value of the shrinkage parameter λ that minimize the mean square error of the estimates. The application of this

29 See, for example, Weighting for Unequal Pi. Kish, L. (1992). Journal of Official Statistics, Vol.8, No.2, Potter, F.J. (1988) Survey Procedures to Control Extreme Sampling Weights, Proceedings of the Section on Survey Research Methods, American Statistical Association and Liu, B., Ferraro, D., Wilson, E., and Brick, M. (2004). Trimming Extreme Weights in Household Surveys. ASA Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association, 2004

methodology for the optimal value $\lambda = 0.86$ reduces the extreme weights without any relevant distortion of the original weights or the estimations.

The inclusion of a shrinkage methodology is suggested by the consultant as an alternative to the ad-hoc current trimming rule applied to 3rd EQLS, which affects to any weight beyond the arbitrary values of 0.3 and 3.

The codes in R to implemented both the calibration algorithm with non-negativity constrains and the shrinkage method are included in a companion file.

Chapter 5: Conclusions and recommendations

After the implementation of the calibration exercise described in these report, the three following main conclusions could be established:

- ***Robustness with respect to the weighting strategy***. As discussed in Sections 3 and 4, the impact of the application of the different weighting scenarios does not generate important changes in the estimated distribution of the analysed variables. This robustness is a very positive feature of 3rd EQLS and, specifically, of its current sampling and calibration methodologies.

- ***Scenario 4 could be considered as the optimal calibration scenario***. As discussed in section 3, Scenario 4 exhibits a performance that is quite similar to that of to the other five scenarios in terms of sampling error and distortion of the design weights. However, estimations under Scenario 4 have a slightly lower bias. This scenario applies the GREG methodology to compute the calibration weights and considers a larger set of variables than that of the current calibration of 3rd EQLS. Specifically, beyond of the current calibration variables (number of households by household size at national level, as well as population by age cross with gender), Scenario 4 includes both the annual activity status at a country level (obtained from EU-SILC) and Education at NUTS 2 disaggregation level (obtained from LFS). The sources for these additional variables are within the ESS and present the required timeliness to be used in the calibration of EQLS. As discussed in Subsection 2.4, EU-SILC is preferred to LFS as statistical source for the Activity Status due to the coherence in the definitions – based in self-assignment - of the Activity Status in EU-SILC and EQLS.

- The addition of **non-negativity constrains** in the GREG algorithm and the application of a **shrinkage methodology** are useful procedures for the treatment of extreme weights. They allow for the elimination of potential negative weights arising in scenario 4 and reduce the extreme differences with the original design weight without introducing significant distortions in the estimates.

The analysis of the current weighting strategy of EQLS and the alternative calibration scenarios supports a series of recommendations that could be applied in the weighting of 4th EQLS:

- *Application of Generalized Regression Estimation (GREG) methods* to compute the weights is recommended instead of the most common approach of Iterative Proportional Fitting (IPF). In EQLS, the latter methodology seems to generate a larger distortion of the design weights, without a relevant reduction of the bias. Moreover, the potential problems that could be generated by the presence of negative weights is not relevant and can be easily corrected as discussed in Section 4.

- *Substitution of the current trimming strategy by a shrinkage method,* where the value of the shrinkage parameter computed to minimize the mean square error of the estimates.

- *Self-weightiness need not be considered as a requirement in EQLS sampling.* 3rd EQLS sampling is designed in order to achieve the objective of self-weightiness. This requirement is quite difficult to fulfil (see the discussion in Section 1) and has the consequence of an assignment of too small sample sizes to the smallest subpopulations. Since the total sampling sizes per country in EQLS are not very large, this fact may have an implication on both sampling errors and calibration procedures. As it can be seen in other social surveys, such as EsocS, self-weightiness is not a requirement itself and sampling designs with different design weights for different countries could be considered in 4th EQLS. In this case, design weights – and the information to replicate its computation - should be also disseminated.

- *Specification of the design weights.* Independently of the final decision on considering or not self-weighted samples for 4th EQLS, it is necessary to record and disseminate all the information required for a detailed computation of the design weights.

- *Calibration by household size.* Calibration using household distribution by household size could be applied in future waves of EQLS, due to the lack of availability of updated statistics of population by household size. As discussed in the report, a higher degree in the harmonisation of the statistical sources to obtain the corresponding reference marginal distributions is recommended. EU-SILC provides with a reliable source for these distributions.

- *Dissemination of the design and calibrated population weights.* The current strategy implemented by Eurofound of dissemination of calibrated weights differs from other alternatives applied in other social surveys. For instance, ESocS computes and disseminates population size weights, meanwhile the approach followed by Eurofound consisting is focused in the dissemination of calibrated weights for the most standard combinations of countries. ESocS and EQLS practices could be integrated in 4th EQLS, with a computation and dissemination of (1) weights for some basic aggregation of countries (as in 3rd EQLS) and (2) design, calibrated and population size weights (as in ESocS).

- *Harmonisation of the geo-codifications.* There are some other variables that could obviously play a role in calibration – such as the degree of urbanisation – and are related to geographical information (i. e. the municipality). To develop all the potential of such calibration variables, it is required a harmonisation of the geo-codifications used in EQLS with those in the ESS, specifically Eurostat's codifications applied in EU-SILC and LFS.

- *The role of Census data in EQLS calibration strategy* should consider the fundamental trade-off between accuracy versus timeliness and coherence that defines this type of data. In other words, the use of census data in calibration may improve the accuracy of the estimates, provided that the reference date of the census is close to the reference data of the corresponding EQLS wave. However, this increase in accuracy could be cancelled by the existence of any time lag - even a short one for very dynamic variables such as employment level – between both reference dates. In this latter case, the use of more recent sampling-based statistical operation could be justified. Finally, the use of census data to calibrate some waves of EQLS and sampling-based data to calibrate other ones could have an impact on the time coherence of the data. For all these reasons, the use of census data to calibrate the 'dissemination' estimates of EQLS should only be considered under a staged approach. A first calibration stage using data from EU-SILC and LFS would provide with a series of estimations that are coherent along the different waves of EQLS. If more recent or accurate data were available in the future, Eurofound could update and disseminate new weights using the new data. To this end, Eurofound should develop an in-house expertise to apply calibrations methods that can take advantage of the R programmes developed for this project.

  On the other hand, census data provide with an exceptional opportunity to check the accuracy of the estimates obtained with sampling-based data, in these occasions where the timing of the census and the EQLS wave coincides. In such cases, a comparison exercise considering both the EQLS actual weighting strategy and a calibration using census data would be a relevant analysis within the quality assessment of EQLS.

- *Recalibration of 1st, 2nd and 3rd EQLS.* If any relevant modification in the weighting is considered for 4th EQLS and for the sake of time coherence, the dissemination of the results of the recalibration of the previous waves using the same methodology is highly recommended.

Annex 1: Coefficient of variation of the selected variables by country

Table A.1.1. Mean of the coefficients of variation.

	IPF: Edu + Emp	IPF: Emp	IPF: Edu	GREG: Edu + Emp	GREG: Emp	GREG: Edu
Austria	0.122	0.123	0.122	0.121	0.121	0.117
Belgium	0.112	0.112	0.112	0.112	0.112	0.111
Bulgaria	0.128	0.127	0.128	0.128	0.127	0.125
Cyprus	0.123	0.120	0.123	0.123	0.120	0.122
Czech Republic	0.128	0.126	0.128	0.126	0.125	0.119
Denmark	0.140	0.139	0.140	0.139	0.138	0.136
Estonia	0.123	0.122	0.123	0.121	0.121	0.115
Finland	0.157	0.157	0.157	0.153	0.153	0.151
France	0.069	0.069	0.069	0.069	0.069	0.067
Germany	0.062	0.062	0.062	0.062	0.062	0.059
Greece	0.112	0.112	0.112	0.112	0.111	0.111
Hungary	0.126	0.126	0.126	0.126	0.126	0.126
Italy	0.085	0.085	0.085	0.085	0.085	0.083
Latvia	0.119	0.119	0.119	0.119	0.118	0.118
Lithuania	0.113	0.112	0.113	0.112	0.111	0.105
Luxembourg	0.112	0.112	0.112	0.112	0.112	0.111
Netherlands	0.118	0.117	0.118	0.118	0.117	0.117
Poland	0.071	0.071	0.071	0.071	0.071	0.071
Portugal	0.119	0.118	0.119	0.119	0.118	0.119
Romania	0.097	0.095	0.097	0.097	0.095	0.095
Slovakia	0.129	0.128	0.129	0.129	0.128	0.129
Slovenia	0.118	0.117	0.118	0.118	0.117	0.117
Spain	0.091	0.091	0.091	0.091	0.091	0.085
Sweden	0.124	0.123	0.124	0.124	0.122	0.121
UK	0.097	0.096	0.097	0.094	0.093	0.079

Table A.1.2. Standard deviation of the coefficients of variation.

	IPF: Edu + Emp	IPF: Emp	IPF: Edu	GREG: Edu + Emp	GREG: Emp	GREG: Edu
Austria	0.122	0.123	0.122	0.121	0.121	0.117
Belgium	0.112	0.112	0.112	0.112	0.112	0.111
Bulgaria	0.128	0.127	0.128	0.128	0.127	0.125
Cyprus	0.123	0.120	0.123	0.123	0.120	0.122
Czech Republic	0.128	0.126	0.128	0.126	0.125	0.119
Denmark	0.140	0.139	0.140	0.139	0.138	0.136
Estonia	0.123	0.122	0.123	0.121	0.121	0.115
Finland	0.157	0.157	0.157	0.153	0.153	0.151
France	0.069	0.069	0.069	0.069	0.069	0.067
Germany	0.062	0.062	0.062	0.062	0.062	0.059
Greece	0.112	0.112	0.112	0.112	0.111	0.111
Hungary	0.126	0.126	0.126	0.126	0.126	0.126
Italy	0.085	0.085	0.085	0.085	0.085	0.083
Latvia	0.119	0.119	0.119	0.119	0.118	0.118
Lithuania	0.113	0.112	0.113	0.112	0.111	0.105
Luxembourg	0.112	0.112	0.112	0.112	0.112	0.111
Netherlands	0.118	0.117	0.118	0.118	0.117	0.117
Poland	0.071	0.071	0.071	0.071	0.071	0.071
Portugal	0.119	0.118	0.119	0.119	0.118	0.119
Romania	0.097	0.095	0.097	0.097	0.095	0.095
Slovakia	0.129	0.128	0.129	0.129	0.128	0.129
Slovenia	0.118	0.117	0.118	0.118	0.117	0.117
Spain	0.091	0.091	0.091	0.091	0.091	0.085
Sweden	0.124	0.123	0.124	0.124	0.122	0.121
UK	0.097	0.096	0.097	0.094	0.093	0.079

Table A.1.3. 5% percentiles of the coefficients of variation.

	IPF: Edu + Emp	IPF: Emp	IPF: Edu	GREG: Edu + Emp	GREG: Emp	GREG: Edu
Austria	0.024	0.024	0.024	0.024	0.024	0.025
Belgium	0.031	0.031	0.031	0.031	0.031	0.031
Bulgaria	0.040	0.040	0.040	0.040	0.040	0.041
Cyprus	0.035	0.036	0.035	0.035	0.035	0.035
Czech Republic	0.038	0.036	0.038	0.038	0.037	0.037
Denmark	0.015	0.015	0.015	0.014	0.014	0.013
Estonia	0.032	0.031	0.032	0.030	0.030	0.029
Finland	0.017	0.017	0.017	0.016	0.016	0.016
France	0.013	0.013	0.013	0.013	0.013	0.013
Germany	0.017	0.018	0.017	0.017	0.017	0.016
Greece	0.030	0.030	0.030	0.030	0.030	0.029
Hungary	0.027	0.028	0.027	0.027	0.027	0.027
Italy	0.023	0.023	0.023	0.023	0.023	0.023
Latvia	0.026	0.026	0.026	0.027	0.027	0.027
Lithuania	0.029	0.029	0.029	0.029	0.029	0.027
Luxembourg	0.024	0.024	0.024	0.024	0.024	0.024
Netherlands	0.014	0.013	0.014	0.014	0.013	0.014
Poland	0.019	0.019	0.019	0.019	0.019	0.019
Portugal	0.021	0.021	0.021	0.021	0.021	0.022
Romania	0.031	0.028	0.031	0.030	0.028	0.031
Slovakia	0.031	0.031	0.031	0.031	0.030	0.030
Slovenia	0.018	0.019	0.018	0.018	0.019	0.018
Spain	0.019	0.019	0.019	0.019	0.019	0.018
Sweden	0.013	0.013	0.013	0.013	0.013	0.013
UK	0.015	0.015	0.015	0.015	0.015	0.013

Table A.1.4. 95% percentiles of the coefficients of variation.

	IPF: Edu + Emp	IPF: Emp	IPF: Edu	GREG: Edu + Emp	GREG: Emp	GREG: Edu
Austria	0.425	0.426	0.425	0.424	0.424	0.413
Belgium	0.389	0.381	0.389	0.387	0.383	0.404
Bulgaria	0.670	0.652	0.670	0.672	0.662	0.642
Cyprus	0.771	0.713	0.771	0.780	0.711	0.792
Czech Republic	0.775	0.743	0.775	0.761	0.739	0.729
Denmark	0.675	0.684	0.675	0.701	0.715	0.736
Estonia	0.420	0.419	0.420	0.418	0.417	0.401
Finland	0.707	0.707	0.707	0.710	0.709	0.710
France	0.175	0.177	0.175	0.172	0.171	0.171
Germany	0.182	0.188	0.182	0.186	0.189	0.174
Greece	0.387	0.390	0.387	0.389	0.391	0.392
Hungary	1003	1003	1003	1003	1003	1003
Italy	0.247	0.250	0.247	0.236	0.239	0.228
Latvia	0.392	0.391	0.392	0.398	0.397	0.409
Lithuania	0.512	0.517	0.512	0.516	0.519	0.433
Luxembourg	0.344	0.345	0.344	0.346	0.347	0.345
Netherlands	0.549	0.563	0.549	0.555	0.564	0.574
Poland	0.340	0.345	0.340	0.346	0.350	0.329
Portugal	0.330	0.328	0.330	0.334	0.324	0.332
Romania	0.570	0.591	0.570	0.614	0.571	0.531
Slovakia	0.453	0.435	0.453	0.459	0.442	0.475
Slovenia	0.598	0.584	0.598	0.596	0.583	0.596
Spain	0.332	0.327	0.332	0.348	0.338	0.307
Sweden	0.537	0.542	0.537	0.541	0.540	0.529
UK	0.323	0.309	0.323	0.270	0.269	0.228

Annex 2: Recalibration exercise with 2011 Census data

This second annex to the final report presents the results of the application of calibration methodology described for Scenario 4 with the 2011 census data. Specifically, it presents a summary of the distortion of the design weights and the prevalence of extreme weights when census data are applied in the recalibration process.

As agreed with Eurofound, the calibration exercise was carried out for those countries (1) whose 2011 Census data were available on October 10th 2014 and (2) the percentage of missing value in the calibration variables Education Level and Activity Status were larger than 5%. Data were not available for HR, IS, KO, ME, MK, SR and TR. Moreover, applying the latter criterion, BE (9.6% of missing values in education), CZ (33.9% of missing values in activity), LU (19.1% of missing values in education), PL (42.1% of missing values in activity) and SK (16.4% of missing values in activity) were excluded of the analysis.

The exercise was then implemented for 22 countries, specifically Austria, Bulgaria, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Netherlands, Portugal, Romania, Slovenia, Spain, Sweden and UK.

Distortion of the original weights

As described in Table 3.1 of the final report, the distortion of the design weights can be measured through the distribution of the ratio of the recalibrated weights (using census data) over the original design weights. The quartiles of such a distribution are shown in Table A.1 at an aggregate level and in Table A.2 at a country level. At an aggregate level, these values are quite similar to the quartiles of the distribution of the ratios with non-census data[30]. For instance, the median before treatment of extreme values is the same when using census and non-census data (0.98) and the difference after shrinkage in only 0.01 (0.97 with census data and 0.98 with non-census data). The interquartile intervals are also very similar.

Table A.1. Quartiles of the ratios of design weights and calibrated weights using census and non-census data.

Weight		P25	Median	P75
Census	Before trimming and shrinkage weight	0.66	0.98	1.32
	Trimmed and shrunken weight	0.65	0.97	1.32
Non-census	Before trimming and shrinkage weight	0.68	0.98	1.26
	Trimmed and shrunken weight	0.72	0.98	1.30

[30] See Table 4.3 in the final report.

Table A.2.Quartiles, minimum and maximum of the ratios of design weights and calibrated weights using census data by country.

Weight		Minimum	P25	Median	P75	Maximum
Total	**Before trimming and shrinkage weight**	**-0.80**	**0.66**	**0.98**	**1.32**	**3.92**
	Trimmed and shrunken weight	**0.10**	**0.65**	**0.97**	**1.32**	**4.01**
Austria	Before trimming and shrinkage weight	-0.31	0.56	0.96	1.45	3.22
	Trimmed and shrunken weight	0.10	0.54	0.96	1.45	3.25
Bulgaria	Before trimming and shrinkage weight	-0.80	0.61	0.96	1.41	3.92
	Trimmed and shrunken weight	0.10	0.59	0.95	1.42	4.01
Cyprus	Before trimming and shrinkage weight	-0.64	0.75	1.13	1.55	3.34
	Trimmed and shrunken weight	0.10	0.73	1.11	1.54	3.42
Denmark	Before trimming and shrinkage weight	-0.48	0.58	0.94	1.47	3.79
	Trimmed and shrunken weight	0.10	0.54	0.95	1.48	3.87
Estonia	Before trimming and shrinkage weight	-0.36	0.68	1.04	1.42	2.46
	Trimmed and shrunken weight	0.10	0.65	1.03	1.43	2.49
Finland	Before trimming and shrinkage weight	-0.78	0.40	0.97	1.64	3.47
	Trimmed and shrunken weight	0.10	0.13	0.83	1.69	3.89
France	Before trimming and shrinkage weight	0.01	0.69	1.03	1.34	2.47
	Trimmed and shrunken weight	0.10	0.69	1.03	1.34	2.46
Germany	Before trimming and shrinkage weight	-0.69	0.62	1.01	1.38	2.81
	Trimmed and shrunken weight	0.10	0.59	0.99	1.40	2.91
Greece	Before trimming and shrinkage weight	0.12	0.71	0.94	1.14	1.63
	Trimmed and shrunken weight	0.13	0.71	0.94	1.14	1.63
Hungary	Before trimming and shrinkage weight	0.05	0.73	1.02	1.22	2.03
	Trimmed and shrunken weight	0.10	0.73	1.02	1.21	2.03
Ireland	Before trimming and shrinkage weight	0.03	0.75	1.02	1.32	2.11
	Trimmed and shrunken weight	0.10	0.75	1.02	1.32	2.10
Italy	Before trimming and shrinkage weight	0.01	0.70	1.02	1.40	3.19
	Trimmed and shrunken weight	0.10	0.70	1.02	1.40	3.19
Latvia	Before trimming and shrinkage weight	0.26	0.74	0.87	1.05	1.74
	Trimmed and shrunken weight	0.26	0.74	0.87	1.05	1.73
Lithuania	Before trimming and shrinkage weight	0.17	0.58	0.97	1.19	2.00
	Trimmed and shrunken weight	0.18	0.58	0.97	1.19	1.99
Malta	Before trimming and shrinkage weight	0.11	0.71	0.91	1.15	2.98
	Trimmed and shrunken weight	0.11	0.72	0.91	1.15	2.97
Netherlands	Before trimming and shrinkage weight	-0.19	0.75	1.03	1.29	2.81
	Trimmed and shrunken weight	0.10	0.75	1.03	1.29	2.81
Portugal	Before trimming and shrinkage weight	0.19	0.72	0.94	1.16	1.94
	Trimmed and shrunken weight	0.19	0.72	0.94	1.16	1.93
Romania	Before trimming and shrinkage weight	-0.26	0.58	1.03	1.29	2.55
	Trimmed and shrunken weight	0.10	0.58	1.03	1.28	2.56
Slovenia	Before trimming and shrinkage weight	-0.11	0.79	0.96	1.21	2.08

	Weight	Minimum	P25	Median	P75	Maximum
	Trimmed and shrunken weight	0.10	0.79	0.96	1.21	2.08
Spain	Before trimming and shrinkage weight	-0.37	0.71	1.04	1.39	2.54
	Trimmed and shrunken weight	0.10	0.69	1.03	1.39	2.59
Sweden	Before trimming and shrinkage weight	-0.34	0.57	0.96	1.44	2.52
	Trimmed and shrunken weight	0.10	0.56	0.96	1.44	2.53
UK	Before trimming and shrinkage weight	-0.45	0.52	0.87	1.30	3.49
	Trimmed and shrunken weight	0.10	0.51	0.86	1.31	3.51

Extreme values

Table A.3 presents a comparison of the prevalence of extreme weights obtained with census and non-census data. The percentage of negative weights at an aggregate level is 1.6%, similar to the value found when using non-census data, given by 1.4%. However, the presence of extreme weights is less relevant when using census data, specifically for those weights lower than 0.4 before shrinkage, which represents 8.3% of the cases with non-census and only 3.6% of the cases with census data. In both cases, there are no weights larger than 4. On the other hand, the prevalence of the smallest weights (lower than 0.3) is higher with census data and shrinkage.

Table A.3. Extreme weights using census and non-census data.

	Weight	Negative	<0.3	<0.4	>3	>4
Census	Before trimming and shrinkage weight	1.6%	4.9%	3.6%	0.2%	0.0%
	Trimmed and shrunken weight	0.0%	7.6%	3.5%	0.3%	0.0%
Non-census	Before trimming and shrinkage weight	1.4%	5.5%	8.3%	0.4%	0.0%
	Trimmed and shrunken weight	0.0%	4.1%	6.1%	0.2%	0.0%

The disaggregation of the prevalence of extreme values by country is presented in Table A.4. A comparison of this table with the corresponding one for non-census data (Table 4.4 in the final report) shows that the high percentage of negative residuals for UK with non-census data disappears when census data are considered. However, the prevalence of negative weights in Finland (the other country with high percentage of negative weights with non-census data) increases until a 13.6% when using census data.

Figure A.4. Extreme weights using census data by country.

Weight		Negative	<0.3	<0.4	>3	>4
Total	**Before trimming and shrinkage weight**	**1.6%**	**4.9%**	**3.6%**	**0.2%**	**0.0%**
	Trimmed and shrunken weight	**0.0%**	**7.6%**	**3.5%**	**0.3%**	**0.0%**
Austria	Before trimming and shrinkage weight	1.7%	7.9%	6.8%	0.5%	0.0%
	Trimmed and shrunken weight	0.0%	10.2%	6.8%	0.5%	0.0%
Bulgaria	Before trimming and shrinkage weight	2.2%	7.7%	2.9%	0.7%	0.0%
	Trimmed and shrunken weight	0.0%	10.3%	5.0%	0.9%	0.1%
Cyprus	Before trimming and shrinkage weight	3.3%	3.7%	2.1%	0.5%	0.0%
	Trimmed and shrunken weight	0.0%	9.0%	2.2%	0.5%	0.0%
Denmark	Before trimming and shrinkage weight	3.3%	10.0%	3.3%	0.9%	0.0%
	Trimmed and shrunken weight	0.0%	14.3%	3.7%	0.9%	0.0%
Estonia	Before trimming and shrinkage weight	1.2%	8.2%	3.0%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	12.0%	0.5%	0.0%	0.0%
Finland	Before trimming and shrinkage weight	13.6%	7.4%	4.0%	0.8%	0.0%
	Trimmed and shrunken weight	0.0%	31.8%	3.7%	2.5%	0.0%
France	Before trimming and shrinkage weight	0.0%	1.4%	2.2%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	1.3%	2.2%	0.0%	0.0%
Germany	Before trimming and shrinkage weight	3.6%	5.7%	4.6%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	12.2%	3.3%	0.0%	0.0%
Greece	Before trimming and shrinkage weight	0.0%	1.4%	1.4%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	1.4%	1.4%	0.0%	0.0%
Hungary	Before trimming and shrinkage weight	0.0%	0.8%	1.5%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	0.8%	1.2%	0.0%	0.0%
Ireland	Before trimming and shrinkage weight	0.0%	1.7%	2.5%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	1.7%	2.5%	0.0%	0.0%
Italy	Before trimming and shrinkage weight	0.0%	4.4%	4.2%	0.1%	0.0%
	Trimmed and shrunken weight	0.0%	4.4%	4.2%	0.1%	0.0%
Latvia	Before trimming and shrinkage weight	0.0%	0.2%	2.3%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	0.2%	2.3%	0.0%	0.0%
Lithuania	Before trimming and shrinkage weight	0.0%	3.6%	2.7%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	3.6%	1.7%	0.0%	0.0%
Malta	Before trimming and shrinkage weight	0.0%	1.0%	5.2%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	1.0%	5.2%	0.0%	0.0%
Netherlands	Before trimming and shrinkage weight	0.3%	3.1%	2.4%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	3.6%	2.6%	0.0%	0.0%
Portugal	Before trimming and shrinkage weight	0.0%	0.8%	1.5%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	0.8%	1.5%	0.0%	0.0%
Romania	Before trimming and shrinkage weight	0.6%	10.2%	5.3%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	10.8%	6.0%	0.0%	0.0%
Slovenia	Before trimming and shrinkage weight	0.2%	1.5%	4.4%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	2.0%	4.1%	0.0%	0.0%

	Weight	Negative	<0.3	<0.4	>3	>4
Spain	Before trimming and shrinkage weight	2.2%	7.5%	2.3%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	10.6%	2.0%	0.0%	0.0%
Sweden	Before trimming and shrinkage weight	1.4%	6.1%	5.7%	0.0%	0.0%
	Trimmed and shrunken weight	0.0%	8.1%	6.1%	0.0%	0.0%
UK	Before trimming and shrinkage weight	1.8%	9.3%	6.0%	1.0%	0.0%
	Trimmed and shrunken weight	0.0%	11.9%	5.9%	1.1%	0.0%