ORIGINAL PAPER

Using poverty maps to improve the design of household surveys: the evidence from Tunisia

Gianni Betti[1] · Vasco Molini[2] 🔟 · Dan Pavelesku[3]

Abstract

In this paper we aim to propose a new method for improving the design effect of household surveys based on a two-stage design in which the first stage clusters, or Primary Selection Units (PSUs), are stratified along administrative boundaries. Improvement of the design effect can result in more precise survey estimates (smaller standard errors and confidence intervals) or in the reduction of the necessary sample size, i.e. a reduction in the budget needed for a survey. The proposed method is based on the availability of a previously conducted poverty maps, i.e. spatial descriptions of the distribution of per capita consumption expenditures, that are finely disaggregated in small geographic units, such as cities, municipalities, districts or other administrative partitions of a country that are directly linked to PSUs. Such information is then used to select PSUs with systematic sampling by introducing further *implicit stratification* in the survey design, so as to maximise the improvement of the design effect. Since per capita consumption expenditures estimated at PSU level from the poverty mapping are affected by (small) standard errors, in the paper we also perform a simulation study in order to take into account this addition variability.

✉ Vasco Molini
  vmolini@worldbank.org

  Gianni Betti
  gianni.betti@unisi.it

  Dan Pavelesku
  dpavelesku@worldbank.org

[1]   University of Siena, Siena, Italy

[2]   The World Bank, Rome, Italy

[3]   The World Bank, Algiers, Algeria

🙋 Springer

1 Introduction

The proposed methodology has been implemented for the new 2021 Household Budget Survey (EBCNV 2021) in Tunisia,[1] conducted under a cooperation project funded by the World Bank while the underlying poverty mapping is based on the 2015 Household Budget Survey (EBCNV 2015) and the 2014 Population and Housing Census (RGPH 2014).[2]

To establish trust in the measurement of poverty and inequality, statistical offices have to regularly assess and continuously improve the quality of their processes as well as the accuracy of their data. This should be one of the prerequisites in order to prevent misguided policies especially when such policies are based on data from household budget surveys.

During the last decades, national statistical offices worldwide have started to develop reports describing the quality criteria adopted in the surveys and explaining any instances in which these criteria could not be met. One common method for reporting the precision of poverty measures is to include sampling variability or standard errors and, consequently, the confidence intervals of such estimates.

This practice is needed much more when poverty measures are disaggregated at smaller territorial levels or for some sub-groups of the population. Recently, the United Nations Economic Commission for Europe (UNECE) has addressed these issues in a series of conferences resulting in a useful guide for data disaggregation (UNECE 2020).

While in the past, many national sample surveys were conducted with the aim of getting estimates that were significant at national level only nowadays it is becoming more and more important to first define the exact type of *domains* for which national statistical offices intend to calculate, and then publish poverty measures. This is necessary because when such measures were disaggregated at lower territorial levels or for other *domains*, often such estimates ended up being non-significant according to standard errors calculated *ex-post*. The *ex-ante* definition of sub-sample sizes in particular, and sample designs in general, therefore, becomes crucial in order to have statistically significant estimates.

There is obviously a trade-off between more precise *domain* estimates and the total cost of a sample survey. For this reason, several statistical offices have a strong incentive to plan and implement highly efficient sampling designs. Indeed, an efficient sample design is a prerequisite for any sample survey, and it plays a fundamental role in household budget surveys. As a matter of fact, improvements in the design (called 'design effect' or simply *deff*) can result in more precise survey estimates (smaller standard errors and confidence intervals) or in the reduction of the necessary sample size, therefore, a reduction in the budget needed for the surveys.

This paper aims to propose a new method for increasing the efficiency of two-stage household sample surveys by incorporating information from previous

---

[1] *Enquête nationale sur le Budget, la Consommation et le Niveau de vie des ménages* (EBCNV) was first planned to start in 2020, but due to Covid-19 pandemic was postponed to 2021 year.

[2] *Recensement Général de la Population et de l'Habitat* 2014.

small area estimates of per capita consumption expenditures in the selection process of Primary Selection Units (PSUs). These estimates in small areas are spatial descriptions of the distribution of expenditures finely disaggregated in small geographic units, such as cities, municipalities, districts, or other administrative partitions within a country.

To improve the sample design in a multi-stage scheme, these maps should satisfy at least two key properties: (i) the small area should be directly linked to the PSUs of the design; (ii) standard errors (precision) of such small area estimates should be relatively small and they should be easily calculated by the National Statistical Office conducting the household survey.

For the above reasons, the methodological originality of this contribution is twofold. First, the new method takes into account the theory of small area estimation (SAE); in particular, in the context of poverty estimation, it considers methods developed in World Bank group by Elbers, Lanjouw and Lanjouw (ELL, 2002, 2003), later refined by Tarozzi and Deaton (2009) and Van der Weide (2014); this for a notable improvement in the design effect of household surveys and specifically for Household Budget Surveys (HBSs). The proposed method, however, is very versatile and can be implemented also using alternative approaches to consumption expenditures estimation at 'small area level' (as defined in Rao 2003): in particular, the Empirical Best (EB) of Molina and Rao (2010), and the more recent proposal CensusEB of Correa et al. (2012), later developed in Molina (2019). The present paper is based on CensusEB estimator found to be superior in terms of bias and mean square error (MSE) to ELL and Molina and Rao's Empirical Bayes prediction estimates (Corral et al. 2020, 2021).

Second, the new method described in this paper implements, for the first time in a developing country, the recent Jackknife Repeated Replication (JRR) method, from Verma and Betti (2011), for estimating the *ex-post* standard errors of consumption expenditures measures from HBSs, that is, *complex measures* from *complex surveys*. In this way, it allows the evaluation of the impact of the so-called *implicit stratification* (or *hidden*; Wolter 1985), which is present when PSUs are selected with systematic sampling from one or more lists that are ordered according to some variables correlated with the phenomenon under investigation.

The present paper is composed of five sections. Following this introduction, Sect. 2 presents the theory behind the estimation of standard errors in complex surveys. Attention is paid to the two-stage stratified sampling design, where, in the first stage, selection of PSUs is performed with systematic sampling, i.e. where implicit stratification is present. Since systematic sampling can be seen as a random selection of a single PSU from each implicit stratum, it is not possible to derive a consistent estimator of the sampling variance (Wolter 1985). Therefore, recent replication methods have been developed and implemented for estimating variance of complex measures from complex surveys. Among the most adopted methodologies are bootstrap, Taylor Linearisation, Jackknife Repeated Replications and Balanced Repeated Replications (BRR). On this basis, the second part of Sect. 2 briefly presents the recent methodology of Verma and Betti (2011) for the implementation of a variant of JRR in the case of systematic selection of

PSUs and the possibility of decomposition of the design effect in several components (Kish 1965; Verma 1991; Liu et al. 2002).

The definition of these components is useful for highlighting the proposed implicit stratification; as previously mentioned, we aim to introduce a further implicit stratification by selecting the PSUs with a systematic sampling within each explicit strata. PSUs are ordered by their relative ranking in terms of the per capita consumption expenditure obtained from the poverty mapping approach. Therefore, Sect. 3 is devoted to a brief presentation of the World Bank methodology.

Since, in general, we believe that any new methodological proposal is useful if it can be put into practice, the second part of Sect. 3 presents the implementation of poverty mapping in Tunisia, based on the 2014 Census and the 2015 Household Budget Survey (EBCNV). Moreover, Sect. 4 is devoted to the presentation of the new sample design in the case of the 2021 Household Budget Survey in Tunisia. Three scenarios are taken into account: (i) PSUs are selected with Simple Random Sampling (SRS) within the 24 Governorates (explicit strata) in Tunisia; (ii) PSUs are, first, stratified among the 24 Governorates and then further stratified (implicitly) with systematic sampling by ordering the lists according to some economic and geographical variables, which consequently make the list a geographical serpentine; (iii) PSUs are, first, stratified among the 24 Governorates, and then further stratified (implicitly) with systematic sampling by ordering the lists of PSUs according to the average per-capita consumption expenditure estimated by the poverty mapping method presented in Sect. 3.

This third scenario represents the main original contribution of this paper, and the design effect gain is compared with other scenarios. Such gain in the design effect has an impact on both the estimated variance (for higher precision in the estimates), and in the optimal sample size (i.e. for a smaller budget). Since per capita consumption expenditures estimated at PSU level are affected by (small) standard errors, in the Sect. 4 we also perform a simulation study in order to take into account this additional variability: this is the second original contribution of the paper.

Finally, Sect. 5 concludes the paper; after summarising the main results of the article, the section aims to briefly present some variants of the proposed methodology when Census and/or Household Budget Surveys are not available for recent years in order to perform the poverty mapping exercise.

2 Theory of design effects, variance estimation, resampling techniques

This section aims to present the theory behind the estimation of standard errors in complex surveys such as HBSs. Since the *implicit stratification* proposed is based on capita consumption expenditure performed with the World Bank ELL method (described in further detail in Sect. 3), here the notation adopted is a mixture of the notation for sampling techniques in Wolter (1985) and the notation for poverty mapping in Elbers et al. (2003), as follows.

We have a *population* divided in L strata (l: 1, 2,…L) and clustered in M_l clusters (m: 1, 2,…M_l); in each cluster m there are K_m households (k: 1, 2,…,K_m) where HS_k individuals live (j: 1, 2,…,HS_k). Population size N can therefore be defined as:

$$N = \sum_{l:1}^{L} N_l = \sum_{m:1}^{M} K_m$$

where N_l is the number of households in stratum l. From this population, and for each stratum l (l: 1, 2,…L) we select a *sample* of C_l clusters (Primary Selection Units, PSU; c: 1, 2,…,C_l); within each PSU c we then select h_c households (h: 1, 2,…,h_c), where the number of households in each stratum is denoted by n_l. Then, sample size n can, therefore, be defined as:

$$n = \sum_{l:1}^{L} n_l = \sum_{c:1}^{M} h_c$$

Taking into account a general two-stage sampling design where, in the first stage, clusters (PSUs) are stratified in L strata and are drawn with equal probability; and where, in the second stage, households are drawn from each selected cluster with equal probability, then we can define the estimator of the total population (Wolter 1985 p. 13): $\hat{Y} = \sum_{l:1}^{L} \hat{Y}_l$, where:

$$\hat{Y}_l = \left(N_l/n_l\right) \sum_{c:1}^{C_l} K_c \sum_{h:1}^{H_c} y_{ch}/h_c \tag{1}$$

In some cases, when the size is known, PSUs could be selected with probability proportional to size (PPS): this may increase the efficiency of the sampling scheme, by reducing the design effect defined in Eq. (3) below.

Since the L strata are disjoint sub-sets of the population (they constitute a *partition*), then they are independent of each other, and, therefore, the (true) variance of estimator (1) is $Var\{\hat{Y}\} = \sum_{l:1}^{L} Var\{\hat{Y}_l\}$, where:

$$Var\{\hat{Y}_l\} = N_l^2\left(1 - f_{1l}\right)\left(1/n_l\right)S_{1l}^2 + \left(N_l^2/n_l\right)S_{2l}^2 \tag{2}$$

and where S_{2l}^2 is variance *within* clusters, while S_{1l}^2 is the variance *between* clusters in a certain stratum *l*. Equation (2) will be the starting point for attempting to improve a household survey design.

In some surveys, when samplers are forced to group elements into clusters, they try to reduce the variance (2) by minimising the components *within* clusters S_{2l}^2; however, the construction of clusters cannot be modified easily when planning a household survey such as a HBS, where clusters represent the Census Enumeration Areas (EAs), or other administrative units, such as municipalities or counties. Therefore, efforts have been made to try to reduce the effect of variance *between* clusters S_{1l}^2 in each stratum *l*. In practice, clusters or PSUs are ordered according to some external variables available in the frame list, which are in some way correlated with the phenomenon under investigation Y: this allows to get more homogenous clusters and then it permits a

reduction in variance *between* clusters $S_{1l'}^2$. To make the sampling design more practical, the selection of clusters is performed with systematic sampling of such ordered frame lists. According to Kish (1965) this type of design can be seen as an (*implicit*) stratified sampling, where one single cluster or PSU is selected from each (*implicit*) stratum.

In developing countries, implicit stratification is often introduced in household surveys by sorting clusters in a certain geographical order within each explicit stratum, since frame lists of EAs may not contain other variables correlated with the phenomenon under investigation. In this way, the classical construction of a *serpentine* of clusters, is the most effective method that has been implemented so far for the preparation of the list frame of clusters from which the systematic sampling is conducted. In the present paper we aim to further improve the implicit stratification of a geographic serpentine, by ordering the clusters according to another socio-economic characteristic: a specific description of the living conditions of the clusters included in the so-called *poverty mapping*. This approach will be further described in Sect. 4.

However, it is well-known in the literature that this design of systematic selection of clusters in the first stage makes it impossible to find the exact formula for an unbiased estimator of the variance in (2), which could be renamed as $Var\{\hat{Y}_{sys}\}$, referring to systematic sampling. Wolter (1985) and Cochrane (1977) proposed several methods of approximate estimators of $Var\{\hat{Y}_{sys}\}$, while in the last two decades, replication methods have reached more precise and unbiased estimators, which include Taylor Linearisation (TL), Jackknife Repeated Replications (JRR), Balanced Repeated Replications (BRR), and Bootstrap and Random Groups (RG). Recently, Verma and Betti (2011) have proposed a developed version of JRR, while Arnab et al. (2015) have proposed some more refined methods of variance estimation in the case of systematic sampling in the first stage. They show that such proposed methods and JRR perform better than other methods that are usually adopted by National Statistical Offices in developing countries.

In light of the above, the variance formula (2) is not applicable when estimating complex measures from complex sample designs, but they are very useful as benchmarks for comparison (Kalton et al. 2015). Such comparisons could also be done with reference to Simple Random Sampling (SRS). In fact, Kish (1965) firstly defined the so-called *design effect* as "…the ratio of the variance of any estimate, say, y, obtained from a complex design to the variance of y that would apply with a SRS or unrestricted sample of the same size". In formulas, *Deft²* (*Deft* in terms of standard errors), could be written as:

$$Deft^2 = \frac{Var\{\hat{Y}_{complex}\}}{Var\{\hat{Y}_{SRS}\}} \tag{3}$$

It could be very useful to decompose *Deft²* in several multiplicative parts; the most important ones are: (i) effect of stratification (usually, but not always,[3] smaller

---

[3] Not always in the case of designs for multi-topic household surveys. In these cases, in fact, generally rural/poor areas are oversampled to get enough observations for domain level analysis on these populations – which has the opposite effect (design greater than 1).

than 1); (ii) effect of clustering (greater than 1); (iii) effect of sample weighting adjustment (greater than 1; known as the *Kish effect*).

In Sect. 3, we provide a brief description of the theoretical background of poverty mapping and its implementation in Tunisia from 2014 to 2015.

3 Theory of poverty mapping and its application to Tunisia

There are several methodologies for estimating poverty at a local level; one of the first was developed by Elbers et al. (2002, 2003), and it is often known as the ELL approach. In general, such model-based techniques are defined as Small Area Estimation (SAE) methods (Rao 2003), and may fall in one of the broad categories: (i) area based models (for examples Fay and Herriot 1979) in which the model is estimated on the basis of areas as statistical units; and (ii) unit-level models, where instead the unit of analysis in the model is the individual or the household. Within this second category, Molina and Rao (2010) have innovated with the Empirical Best (EB): in the general literature of SAE, both ELL and EB are nested error models used for small area estimation, which were originally proposed by Battese et al. (1988). Later, the World Bank improved the original ELL method by including EB predictors and incorporating heteroskedasticity and survey weights in order to take into account for complex sampling designs (Van der Weide 2014).

The present paper is based on the CensusEB estimator found to be superior in terms of bias and mean square error (MSE) to ELL and Molina and Rao's Empirical Bayes prediction estimates (Corral et al. 2020, 2021).

The current poverty mapping method imputes consumption expenditure for census households based on a model estimated from the household survey by applying the estimated coefficients of the model to the same variables from the census data. Poverty and inequality statistics for each small area are then calculated with the imputed consumption estimates of census households. The method does not only estimate consumption expenditures means and poverty indices, but also their corresponding standard errors.

This methodology involves three major steps. The first step is to select a set of variables that are common to the census and the household expenditure survey. The subset of variables that are found both in the census and the survey is used to estimate a regression model of per capita consumption using the survey data. In the second step, the set of parameter estimates obtained from the regression model is applied to the same set of variables identified in the census data to obtain predicted per capita consumption for each census household. Finally, based on the estimated level of per capita consumption, estimates of poverty, inequality and other welfare measures, as well as their standard errors, are calculated for any geographical unit with enough households to obtain reliable estimates. The point estimates and standard errors of the welfare indicators are calculated by Monte-Carlo simulations. In each simulation, a set of values for the model parameters are drawn from their estimated distributions, and an estimate of consumption expenditure and poverty rates are obtained.

Table 1 Average per capita consumption estimates by region. *Source*: Our elaborations from INS (2020)

Region	Poverty mapping—Census 2014		
	Average consumption	Standard error	cv
Tunisia	3551	11.5	0.003
Greater Tunis	4978	31.0	0.006
North–East	3365	20.5	0.006
North–West	2564	18.9	0.007
Centre-East	3781	20.0	0.005
Centre-West	2362	15.8	0.007
South–East	2957	22.3	0.008
South–West	2865	26.3	0.009

3.1 Poverty mapping in Tunisia

This section describes the implementation of the method and key findings of small-area estimation in Tunisia. It provides per capita consumption expenditure at the national level, and disaggregated at regional and governorate levels—which constitute official welfare measures—and also disaggregated at the PSU level—which are measures used for implementing the new and original design of the EBCNV 2021. The results in this section are derived from micro-data in the Population and Housing Census (RGPH) 2014 and the National Survey of Budget, Consumption and Standard of Living of Households (EBCNV) of 2015, which has the typical structure of HBSs. Details on the regressors included in the models are reported in INS (2020).

The time lag between 2014–2015 (the performing of the poverty mapping exercise) and 2021 (the use of such poverty mapping) could slightly limit the big potentialities of the proposed approach. We will discuss these limitations and propose two general solutions in the concluding remarks. Unfortunately, we were not able to implement such solutions in this particular case for Tunisia.

Mapping the average per capita consumption in Tunisia enables to show that there is a high concentration of poverty in the North-West and Centre-West landlocked regions, as shown in Table 1. The situation improves moving towards the coastal North-East, Centre-East and the Greater Tunis regions, although there are pockets of relatively low consumptions there as well. Southern regions (South West and South Est) are characterized by diverse grades of poverty.

The performance of the models is tested by comparing the consumption expenditures obtained with the CensusEB method with the EBCNV 2015 direct estimates at the Governorate levels, reported in Table 2. This comparison is feasible because the EBCNV 2015 is representative at both regional and Governorate levels.

The comparative reduction in variance of the estimates based on CensusEB is evident: while sub-sample sizes are decreasing from regions to Governorates, standard errors of direct estimates on the left-side of Table 2 increase exponentially (see Sect. 4 for the methodology adopted), that is, the coefficient of variations are

Table 2 Average per capita consumption estimates by Governorate. *Source*: Our elaborations from EBCNV 2015 and INS (2020)

Governorate	EBCNV 2015			Poverty mapping—Census 2014		
	Average consumption	Standard error	cv	Average consumption	Standard error	cv
Tunis	5810	296.8	0.051	5458	53.2	0.010
Ariana	5461	274.3	0.050	5171	79.9	0.015
Ben Arous	4878	177.3	0.036	4743	45.9	0.010
Manouba	4377	296.8	0.068	3833	47.1	0.012
Nabeul	3919	113.6	0.029	3696	29.2	0.008
Zaghouan	3052	115.4	0.038	3124	54.4	0.017
Bizerte	2868	88.1	0.031	2968	29.6	0.010
Beja	2472	101.9	0.041	2546	36.4	0.014
Jendouba	2943	96.0	0.033	2694	33.3	0.012
Le Kef	2363	69.4	0.029	2317	35.9	0.015
Siliana	2934	135.7	0.046	2628	45.6	0.017
Sousse	3777	119.1	0.032	3737	39.1	0.010
Monastir	5115	208.1	0.041	4114	37.4	0.009
Mahdia	3196	210.1	0.066	2619	36.4	0.014
Sfax	4698.13	133.9	0.029	4135	33.4	0.008
Kairouan	2269	62.5	0.028	2312	22.5	0.010
Kasserine	2543	92.3	0.036	2315	31.9	0.014
Sidi Bouzide	2664	75.7	0.028	2474	28.7	0.012
Gabes	3043	83.4	0.027	2996	37.8	0.013
Mednine	3319	91.1	0.027	2947	32.0	0.011
Tataouine	3539	106.9	0.030	2893	53.2	0.018
Gafsa	3155	106.1	0.034	2889	36.5	0.013
Tozeur	3192	106.4	0.033	3014	62.4	0.021
Kebili	2834	72.0	0.025	2712	45.3	0.017

higher than 4% in one third of Governorates. On the other hand, standard errors of measures calculated according to the CensusEB method show a very low increase. Indeed, coefficient of variations are all less than 2% except in the case of the Governorate of Tozeur.

In the analysis of the measures at lower levels, in particular at the PSU level, direct estimates are affected by too high sampling errors and are, therefore, no longer significative. On the other hand, the CensusEB estimates still show very limited standard errors: it is fundamental to bear in mind that although the PSU-level estimates are not suitable for dissemination and for policy-making, they can still be very useful in informing sample design.

Moreover, even if poverty mapping estimates could be affected by model bias (in Table 2 it seems that such estimates are often lower than EBCNV ones), if this bias

Table 3 Summary statistics on CVs at PSU level by Governorate. *Source*: Our elaborations from INS (2020)

Governorate	1501 PSUs in the sample			Other PSUs		
	Minimum	Median	Maximum	Minimum	Median	Maximum
Tunis	0.031	0.054	0.070	0.021	0.063	0.396
Ariana	0.032	0.051	0.084	0.021	0.060	0.448
Ben Arous	0.030	0.054	0.069	0.017	0.060	0.526
Manouba	0.031	0.053	0.062	0.021	0.058	0.407
Nabeul	0.025	0.054	0.077	0.022	0.061	0.459
Zaghouan	0.045	0.054	0.077	0.026	0.061	0.579
Bizerte	0.026	0.049	0.066	0.027	0.060	0.539
Beja	0.040	0.054	0.079	0.035	0.064	0.500
Jendouba	0.039	0.054	0.069	0.023	0.060	0.610
Le Kef	0.043	0.056	0.069	0.024	0.061	0.453
Siliana	0.042	0.058	0.070	0.027	0.066	0.455
Sousse	0.032	0.059	0.084	0.027	0.066	0.566
Monastir	0.036	0.054	0.072	0.024	0.058	0.379
Mahdia	0.036	0.055	0.080	0.029	0.063	0.591
Sfax	0.031	0.045	0.088	0.022	0.054	0.513
Kairouan	0.031	0.050	0.064	0.026	0.058	0.481
Kasserine	0.031	0.055	0.078	0.021	0.061	0.583
Sidi Bouzide	0.035	0.050	0.064	0.031	0.057	0.513
Gabes	0.034	0.055	0.068	0.021	0.062	0.499
Mednine	0.032	0.055	0.076	0.023	0.064	0.540
Tataouine	0.039	0.055	0.072	0.030	0.062	0.649
Gafsa	0.037	0.056	0.075	0.024	0.063	0.570
Tozeur	0.041	0.054	0.084	0.028	0.059	0.600
Kebili	0.038	0.050	0.074	0.037	0.058	0.428

(underestimate) is systematic and not erratic, this should not have a negative impact on the PSUs ordering.

In order to presents such results, we synthesised measures at the Governorate level—and based on CensusEB—in Table 3. For each Governorate, the table reports the minimum estimated value of the coefficient of variation (i.e. standard error over per capita consumption expenditure), the maximum value of the coefficient of variation, and the median value. The Table 3 reports such statistics disaggregated for the 1,500 PSUs included in the 2015 survey (3.84% of the total) and for those not included (96.16% of the total). Although estimates are systematically better-off for those PSUs included in the sample, we can see that in many cases per capita consumption estimates for any cluster are very precise, with an error of even less than 2% of the estimated mean. In the column "Median" we may conclude that in all Governorates at least 50% of cluster estimates have an error smaller than 10% of the estimated mean. Finally, in the column "Maximum" we can observe that in most of
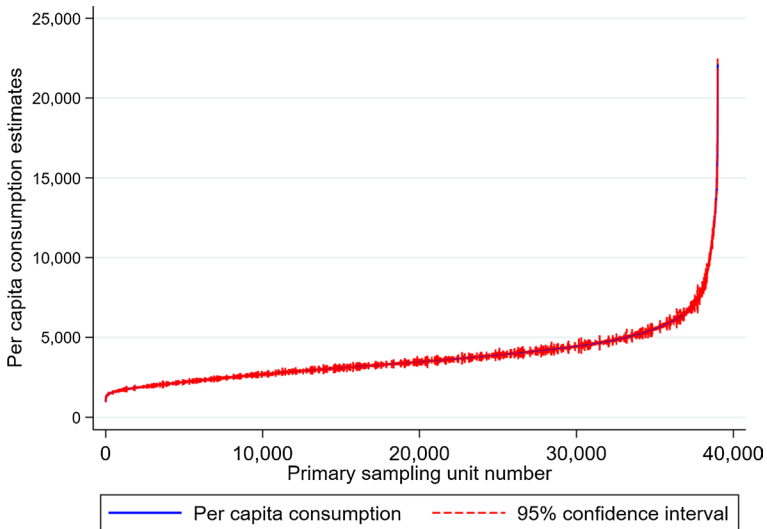
Fig. 1 Estimated per capita consumption (TND per annum) for each PSU (39,001)

cases the error is smaller than 60% of the estimated mean. In other words, there are very few cases (clusters) in which the estimate precision is not high enough and the estimates are not significative.

In order to reinforce the findings of Table 3, we have examined the 95% confidence intervals of per capita consumption at cluster level, reported in Fig. 1. The graph shows that, once PSUs have been ranked by per capita consumption expenditure, confidence intervals among the richer clusters are disjoint with confidence intervals among clusters whose per capita consumption are close to the national average value, and disjoint with confidence intervals among the poorer clusters.

4 Implementation of the new design in Tunisia: EBCNV 2021

In this section, we discuss the implementation of the proposed design in the case of the new 2021 EBCNV in Tunisia. This could easily be extended to any HBS in any country and—with some specific adjustments—to any household survey.

Tunisia has conducted ten rounds of EBCNVs so far, starting in 1966, then in 1974, and from 1980 one has been completed every five years. In the ninth edition, in 2010, the sample size of about 13,400 households allowed the survey results to be representative at regional levels, such as the seven regions reported in Table 1. For the tenth edition, in 2015, INS (2015) aimed to increase the survey sample size to let the survey estimates to be representative at the level of the 24 Governorates. In order to reach this goal, 24 explicit strata were defined, one for each Governorate, and the new sample size was fixed at more than 25,000 households. This sample size was defined on the basis of: (i) the information of the population at that time; (ii) the

explicit stratification of the PSU in the design, and (iii) the expected maximum error of 10% at the Governorate level.

However, the EBCNV 2015 had a more efficient design than that taken into account for selecting the sample size. PSUs were selected with systematic sampling, ordering PSUs by the average household size and prevalent economic status of households, as described in a document from the previous survey (INS, 2010). Moreover, it has been verified *ex-post* that further implicit stratification was also introduced by ordering PSUs, at the minimum, by delegation, thus forming a sort of geographic serpentine. These three factors contributed to better implement the design and the resulting errors became smaller than those expected when planning the design. Taking into account such implicit stratification, Betti and Pavelesku (2019) have estimated the real design effect of the EBCNV 2015 and the relative standard errors for estimates of mean per capita consumption expenditures at Governorate level. These results are those reported in the left side of Table 2.

Since the aim of the EBCNV 2015 was to reach significative estimates at the Governorate level, sub-sample sizes were planned to be as proportional as possible to sub-population sizes. This fact has resulted in homogenous post-stratification or calibration weights and small Kish effects.

During the first planning phases of the new EBCNV 2021, the INS initially took into account three aspects for determining the sample size: (i) the information of the population at that time; (ii) the *explicit stratification* of the PSU in the design, and (iii) the expected maximum error of 10% at the Governorate level. These would have resulted in a sample size of about 29,000 households and a total cost of about 6.9 million Dinars (about 2.3 million US dollars).

Given the comparative benchmark of 6.9 million Dinars and the benchmark of expected maximum error of 10% at the Governorate level, the aim of the new sampling design has been to reduce both costs and maximum errors. Three scenarios have been initially taken into account, with different survey designs in terms of design effect and sample sizes needed to reach specified maximum percentage errors. The three scenarios are as follows:

(a) *Explicit stratification only* PSUs are stratified in the 24 Governorates, and then selected with simple random sampling; this is an unrealistic scenario nowadays, even in household surveys in Africa. However, the expected standard errors are usually calculated hypothesising this scenario so that the implemented sample sizes are overestimated for the purposes of the survey with very high costs.

(b) *Serpentine* PSUs are first stratified in the 24 Governorates, and then further stratified (implicitly) with systematic sampling by ordering the lists according to some economic and geographical variables, which consequently form the list as a geographical serpentine.

(c) *Poverty map* This is the design proposed in the present paper; PSUs are firstly stratified in the 24 Governorates, and then they are further stratified (implicitly) with systematic sampling by ordering the lists of PSUs by average per capita consumption expenditure estimated with the poverty mapping method presented in Sect. 3.
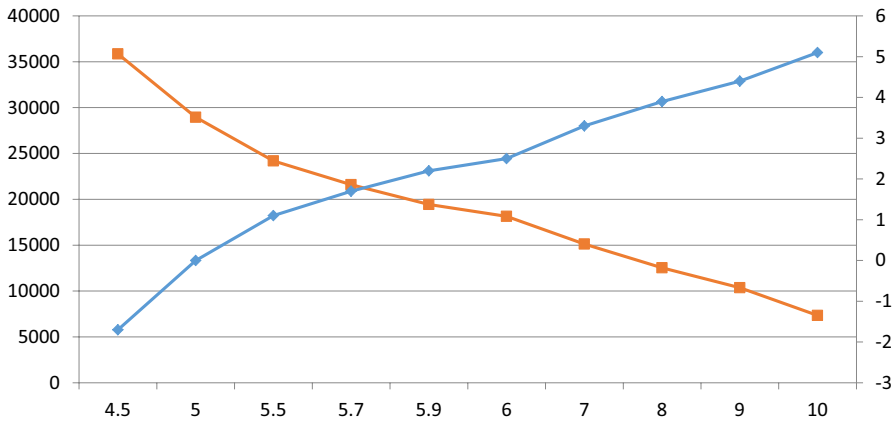
Table 4 Sample sizes in function of expected errors and type of design

	Deft (%)	Explicit stratification 3.71	Serpentine 2.19	Poverty map 1.87 (underestimated)	Poverty map 1.97 (simulation study)
error	4.5	141,315	49,306	35,856	39,897
	5.0	114,073	39,801	28,944	32,206
	5.5	95,345	33,267	24,192	26,919
	5.7	85,129	29,702	21,600	24,035
	5.9	76,616	26,732	19,440	21,631
	6.0	71,509	24,950	18,144	20,189
	7.0	59,591	20,792	15,120	16,824
	8.0	49,375	17,227	12,528	13,940
	9.0	40,862	14,257	10,368	11,537
	10.0	28,944	10,099	7344	8172

On the basis of the new information from the current population in Tunisia, and the design effect of 2.19 estimated on the basis of the EBCNV 2015, Table 4 reports the expected sample sizes to obtain maximum percentage errors reported in the left column. The expected design effect corresponding to scenario a) has been estimated using Eq. (2) in Sect. 2, where Y is the mean per capita consumption expenditure. First of all, we can observe that in order to obtain a percentage error of about 10% a very high sample size is not needed when good designs with implicit stratification are taken into account (such as the classical "Serpentine"). However, this can be verified when very complex methods of variance estimation, usually based on replication techniques, are implemented. The recent JRR version, proposed by Verma and Betti (2011), and implemented to estimate the design effect 2.19 of the "Serpentine" design, is much more easily implemented by National Statistical Offices.

The further use of implicit stratification proposed by the "Poverty map" design is able to further reduce the estimated design effect to 1.87 (also estimated with the JRR method, though that might be an underestimation due to the small variability in the real ranking of the PSU in the implicit stratification). This could lead to a reduction in the needed sample size (and costs) and the desired maximum percentage error.

Based on the (c) column of Table 4, the INS has decided to adopt a final size of 21,600 households selected from 1800 PSUs. This choice, also seen in Fig. 2, is based on two main considerations: (i) this size not only permits getting maximum percentage errors less than 10%, that can also reach an expected value at around 5.7%, which is even below the estimated maximum value in the EBCNV 2015; (ii) thus, the total cost of the survey will be about 5.2 million Tunisian dinars (about 1.7 million US dollars), with a 1.7 million dinar decrease in costs (gain) when compared with the initial estimated budget.

Fig. 2 Sample sizes (left) and gains in cost (right) as function of expected maximum error

4.1 Simulation study

In scenario c) the design effect has been estimated as if the average consumption expenditures estimates at PSU level from the CensusEB exercise were not affected by (small) standard errors. In order to take into account this additional source of variability, we have conducted a simulation study.

In each iteration of the study, we have drawn an average per-capita consumption expenditure for each PSU in the Census, taking into account the available replicates from the poverty mapping exercise; then, we have re-drawn the PSUs with systematic sampling and the same starting points of scenario c). We have performed up to 100 iterations in order to take into account the variability in the selection of the 1800 PSUs in the first stage.

In most cases the same PSUs have been re-drawn: the average rank correlation index at country level has been equal to 0.917. This corresponds to an increase of the inefficiency of about 3.5% and a corresponding design effect of 1.97.

The right-side column of Table 4 reports the relationships between sample size and coefficient of variation of average consumption expenditure in EBCNV 2021 with the expected design effect of 1.97 estimated in the simulation study. The real sample size of about 21,600 interviewed households is able to maintain the final coefficient of variation of survey estimates at National level below 6%, and well below the INS target of 10%.

5 Discussion, further research and concluding remarks

In this paper we have proposed an original methodology for improving the design effect of household surveys in general and Household Budget Surveys in particular. This is carried out by using the results of a poverty mapping exercise implemented

by the method developed by the World Bank. Moreover, we have described how the new JRR method, proposed by Verma and Betti (2011), has been implemented for the first time in a developing country by a National Statistical Office for estimating standard errors for *complex measures* from *complex surveys*.

This discussion should take into account three main issues concerning the implementation of a poverty mapping for improving the design of household surveys: (i) the capacities of National Statistical Offices to perform poverty mappings; (ii) the problem of performing a poverty mapping when the census has been conducted too long ago; (iii) the variability affecting the mean per capita consumption per PSU in the poverty mapping exercise.

Concerning issue (i), the poverty mapping exercise was initially quite an intensive procedure which needed high econometric capacity and, therefore, only highly skilled consultants were able to implement it. In order to apply the ELL approach, the World Bank research team has developed a user-friendly software that can easily be implemented inside National Statistical Offices. PovMap version 2.1, the latest version as of 2015,[4] provides computational solutions to all stages of poverty mapping activities. However, the new methods mentioned earlier are developed only in a Stata module produced by a team of researchers at the World Bank.[5] This software is valid for use when there is a recent census available and a recent Household Budget Survey (or World Bank format of Living Standard Measurement Survey, LSMS) also available. An alternative method for introducing additional implicit stratification, and closely related to poverty mapping, would be ranking PSUs according to the covariates used in the regression model of the poverty map model and available in the Census data set; we believe that ranking PSUs according to average values of such covariates would have implicated subjective choices: *which would have been the hierarchical order of covariates*. In particular, the order of the second, third, fourth covariate and so on would have been relevant only when the first or more important covariate has ties among PSUs, i.e. it is categorical at PSU level (and not at household or individual level). From this point of view, the average consumption expenditure is able to summarise the contribution of all covariates at once, giving implicit importance which is proportional to the estimated parameters in the models in the second step of the poverty mapping.

Concerning issue ii), often the question may be: "What should be done when the poverty map is too old?". In this case there are two possible solutions: ii-a) In some countries or cases, the poverty map may be too old or not updated enough in order to properly implement the methodology proposed in this paper. This may occur primarily because a census is usually conducted every ten years. There are two possible solutions depending on the type of recent household survey that is available. In the case of a recent HBS or LSMS, the proposed methodology could take into account the so-called "updated" or "further-updated" poverty maps that are prepared with recent methodologies developed under World Bank projects. In particular, Dabalen and Ferrè (2008) have shown how to implement the Lemieux (2002) approach to

---

[4] https://www.worldbank.org/en/research/brief/software-for-poverty-mapping.

[5] The Stata package is available here https://github.com/pcorralrodas/SAE-Stata-Package.

link the new HBS with the old HBS, and then update the poverty mapping through the use of the old Census. Moreover, Betti et al. (2013) have demonstrated that this updating is still valid also after several years, that is, when the poverty mapping is further updated with a third, more recently updated HBS. ii-b) Otherwise, if a more recent HBS or LSMS is not available but, instead, there is a recent survey not covering consumption expenditure information, the proposed methodology could be complemented by the two methods proposed by Azzarri et al. (2006), or Dang et al. (2014, 2017).

The third issue covered in this discussion is that the per capita consumption expenditures estimated at PSU level from the poverty mapping exercise are affected by (small) standard errors. Such variability has been taken into account with a simulation study, which has refined the original contribution proposed in this paper. In the context of the poverty map in Tunisia, such variability has generated an increase of standard errors of about 3.5%, and coefficient of variation passing from about 5.7% to about 5.9%, still well below the target set up by the INS at 10%.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Azzarri C, Carletto G, Davis B, Zezza A (2006) Monitoring poverty without consumption data: an application using the Albania panel survey. East Eur Econ 44(1):59–82

Battese GE, Harter RM, Fuller WA (1988) An error-components model for prediction of county crop areas using survey and satellite data. J Am Stat Assoc 83(401):28–36

Betti G, Pavelesku D (2019) Calculation of the design effect and JRR standard errors in the 2015 Household Budget Survey in Tunisia, Report to The World Bank

Betti G, Dabalen A, Ferré C, Neri L (2013) Updating poverty maps between censuses: a case study of Albania, In: Laderchi CR, Savastano S (eds) Poverty and exclusion in the Western Balkans, Economic Studies in Inequality, Social Exclusion and Well-Being 8, Springer, New York

Cochran WG (1977) Sampling techniques, 3rd edn. John Wiley & Sons, New York

Corral P, Molina I, Nguyen MC (2020) Pull your small area estimates up by the bootstraps. World Bank policy research working paper No 9256. The World Bank, Washington DC

Corral P, Kastelic Himlein K, Mcgee KR, Molina I (2021) A map of the poor or a poor map? World Bank policy research working paper series No. 9620. The World Bank, Washington DC

Correa L, Molina I, Rao JNK (2012) Comparison of methods for estimation of poverty indicators in small areas. Unpublished report

Dabalen A, Ferrè C (2008) Updating poverty maps: a case study of Albania. Unpublished working paper, The World Bank, Washington DC

Dang HA, Lanjouw P, Serajuddin U (2014) Updating poverty estimates at frequent intervals in the absence of consumption data: methods and illustration with reference to a middle-income country. World Bank Policy Research Paper No. 7043, The World Bank, Washington DC

Dang HA, Lanjouw P, Serajuddin U (2017) Updating poverty estimates at frequent intervals in the absence of consumption data: methods and illustration with reference to a middle-income country. Oxf Econ Pap 69(4):939–962

Elbers C, Lanjouw JO, Lanjouw P (2002) Micro-level estimation of welfare. World Bank policy research working paper No 2911, The World Bank, Washington DC

Elbers C, Lanjouw JO, Lanjouw P (2003) Micro-level estimation of poverty and inequality. Econometrica 71(1):355–364

Institut National de la Statistique (INS) (2010) Enquête nationale sur le Budget, la Consommation et le Niveau de vie des ménages pour l'année 2010, Tunis

Institut National de la Statistique (INS) (2015) Enquête nationale sur le Budget, la Consommation et le Niveau de vie des ménages pour l'année 2015, Tunis

Institut National de la Statistique (INS) in collaboration with the World Bank (2020) Carte de la pauvreté en Tunisie, Tunis

Kish L (1965) Survey sampling. John Wiley & Sons, New York

Lemieux T (2002) Decomposing changes in wage distributions: a unified approach. Can J Econ 35(4):646–688

Liu J, Iannacchione V, Byron M (2002). Decomposing design effects for stratified sampling. In: Proceedings of the survey research methods section, American Statistical Association, pp 2124–2126

Molina I (2019) Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas, CEPAL: Comisión Ecónomica para América Latina y el Caribe

Molina I, Rao JNK (2010) Small area estimation of poverty indicators. Can J Stat 38(3):369–385

Rao JNK (2003) Small area estimation. Wiley, New York

Tarozzi A, Deaton A (2009) Using census and survey data to estimate poverty and inequality for small areas. Rev Econ Stat 91(4):773–792

UNECE (2020) Poverty Measurement. Guide to Data Disaggregation, United Nations Economic Commission for Europe, Geneva

Van der Weide R (2014) GLS estimation and empirical Bayes prediction for linear mixed models with heteroskedasticity and sampling weights: A background study for the povmap project. World Bank Policy Research Working Paper No 7028, The World Bank, Washington DC

Verma V (1991) Sampling methods, Statistical Institute for Asia and Pacific, Tokyo

Verma V, Betti G (2011) Taylor linearization sampling errors and design effects for poverty measures and other complex statistics. J Appl Stat 38(8):1549–1576

Wolter K (1985) Introduction to variance estimation. Springer-Verlag