Journal de la Société Française de Statistique
Vol. 155 No. 4 (2014)

Estimation in Multiple Frame Surveys: A Simplified and Unified Review using the Multiplicity Approach

Titre: Enquêtes à bases multiples : un examen simplifié et unifié à l'estimation sous l'approche de la multiplicité

Fulvia Mecatti[1] and Avinash C. Singh[2]

Abstract: Multiple frame surveys are useful for reducing cost for given precision constraints, improving coverage (under or over) and dealing with elusive or rare populations for which a direct sampling frame may not exist. Unlike model-based coverage bias adjustments traditionally used for single-frame surveys where domains of units subject to coverage bias are not identificable, multiple frame surveys assume identifiability of such domains, and supplementary sampling frames along with multiplicity adjustments are used to deal with the coverage bias. Point and variance estimation for multiple frame surveys are somewhat challenging because of multiplicity of units due to overlapping frames, and possible duplication of units in the sample. A simple single-frame solution can be used if selected units from the supplementary frame are screened out whenever they are listed in the main frame. However, this may not be desirable in practice because a major portion of the cost is already incurred in contacting the selected unit for the screening information. Despite the practical appeal of multiple frame surveys, they have not been commonly used possibly because of non-standard complex nature and a lack of general understanding of estimation as well as lack of consensus about a preferred methodology among researchers. However, there has been a recent resurgence of interest in multiple frame due to the practical necessity of mitigating increased cost in data collection and use of non-area frames such as cell and landline telephones. In this paper, we provide a simplified and unified review of different existing methods which should help in a better understanding in choosing a suitable method in any application, and promoting more use of multiple frames in practice.

Résumé : Les enquêtes à bases multiples sont utiles afin de réduire les coûts pour une précision donnée ainsi que pour améliorer la (sous ou sur) couverture et pour le traitement des populations difficiles à joindre ou rares pour lesquelles il n'existe pas une base de sondage directe. Contrairement aux ajustements pour le biais de couverture traditionnellement utilisés pour les enquêtes à bases uniques pour lesquelles les sous-groupes d'unités sujets à des biais de couverture ne sont pas identifiables, les enquêtes à bases multiples font l'hypothèse que les sous-groupes d'unités sont identifiables et utilisent des bases de sondage supplémentaires ainsi que des ajustements pour la multiplicité afin de corriger le biais de sous-couverture. L'estimation ponctuelle et l'estimation de la variance présentent un certain défi dû à la multiplicité des unités provenant de bases chevauchantes et au possible problème de duplicata des unités dans l'échantillon. Une solution basée sur une unique base peut être utilisé pourvu que les unités échantillonnées à partir des bases supplémentaires soient dépistées lorsque présente sur la base principale. Cependant, cela n'est peut-être pas souhaitable en pratique car une partie importante du coût est déjà engagée afin de contacter les unités lors de l'étape de dépistage. Malgré l'attrait pratique des sondages à bases multiples, ils n'ont pas été couramment utilisés probablement en raison de leur nature complexe et non-standard et un manque de compréhension générale de l'estimation ainsi que de l'absence de consensus à propos d'une méthodologie préférée parmi les chercheurs. Cependant, il y a eu un regain d'intérêt récent pour les bases multiples en raison de la nécessité pratique d'atténuer l'augmentation des coûts de collecte des données et de l'utilisation des bases non-aréolaires telles que les téléphones cellulaires et les téléphones fixes . Dans cet article, nous présentons une revue simplifiée et unifiée des différentes

[1] Department of Sociology and Social Sciences, University of Milan-Bicocca.
E-mail: fulvia.mecattil@unimib.it

[2] Center for Excellence in Survey Research, NORC at the University of Chicago.
E-mail: Singh-Avi@norc.org

méthodes existantes, qui permettront de mieux comprendre le choix d'une méthode appropriée dans n'importe quelle application, et d'encourager la promotion d'une utilisation de méthodes à bases multiples.

Keywords: Coverage bias, Elusive/rare populations, GMHT-reg class, Horvitz-Thompson estimation, Imperfect frames, Zero functions,
Mots-clés : Bbases imparfaites, Biais de couverture, class GMHT-reg, estimation d'Horvitz-Thompson, populations rares/difficiles à joindre
AMS 2000 subject classifications: 62D05, 62-02

1. Introduction

Conventional survey sampling methodology assumes at least conceptually the availability of a complete sampling frame consisting of a list of all sampling units each identified by a label. However, *perfect* frames are seldom the case in practice. For instance, the sampling frame can suffer from under- or over-coverage or both. Under-coverage occurs when the available frame is incomplete; i.e., it covers only a part of the target population as in the case of a list of only landline phones without cell phone or non-telephone households in a telephone survey. To overcome this problem, the original frame is supplemented with additional frames such that the union of all frames covers the target population. However, this often gives rise to over-coverage because some units may be present in more than one frame. Over-coverage may also occur in a single complete frame when a unit is duplicated (i.e., counted more than once) because of multiple locations, or when a supplementary incomplete frame is used for cost reasons because it is cheaper to sample than the complete frame, or when a direct sampling frame, even *imperfect*, is not available as in network or indirect sampling. In the practice of survey sampling, the sampling frame is never perfect and there is generally a small fraction of unidentifiable units in the target population that are either missed or over-counted. To adjust for the resulting bias due to under- or over-coverage, the method of post-stratification based on coverage bias models is used to adjust sampling weights. However, when the domain of units subject to under- or over-coverage can be identified, it is preferable and more effective to use the method of multiple frames (MF) to correct for coverage bias without relying on models.

Thus, MF surveys are tools for dealing with imperfect frames. In an MF survey a collection of two or more listings of units is used simultaneously for sample selection. Lists may be complete or incomplete, and may be overlapping with unknown amounts of overlap. As mentioned earlier, MF surveys are often suggested for improving coverage of surveys about difficult-to-sample populations such as elusive and hidden populations as well as rare populations for which a single frame might even be non-existent (see Kalton and Anderson (1986) for a good review). For instance Iacan and Dennis described a survey of the homeless population in which eligible individuals were selected from 1) homeless shelters; 2) soup kitchens and 3) street locations, resulting in a three-frame sampling design (Iacan et al., 1993). Sampling individuals at locations corresponds to an MF setup and has been applied in European migration studies (Eurostat, 2000).

MF surveys can be cost effective even if a complete frame exists. For instance, in an agricultural survey, an area frame of the farms under study can be used. However area frames are usually expensive to sample because of travel costs and in-person interviews for data collection. Suppose that an incomplete list of farms located in the same area is available from an independent source; for example, a list of (email) addresses of certified organic farms. This list would be incomplete and overlapping with the area frame - in fact, nested within the area frame but would be cheaper to sample. It follows that an efficient sampling design could comprise two samples, one from each of the two available frames, by under-sampling the costly complete frame and over-sampling the cheaper incomplete one. Such a motivation for reducing cost for a given precision level was the starting point in the justification of multiple frame (MF) surveys when they were first introduced in the sixties (Hartley, 1962, 1974). MF surveys could also offer higher precision than conventional single frame surveys. For instance, in surveying characteristics of complex populations, modular sampling frames can help to better capture differences between subpopulations or domains as in stratified sampling designs except that domains may not be disjoint. In addition, MF surveys offer a more flexible strategy than conventional single frame surveys. It allows for different sampling designs for different frames as well as different modes of data collection; for instance, face-to-face interviews in one frame and email-questionnaires in another, with the goal of a better control on survey costs, coverage, response rates, and ultimately the estimation accuracy. In contemporary applications, the potential of an MF setup appears even more promising and appealing. For example, in a web survey, both the population non-coverage and the possible bias due to self-selection might be reduced by using multiple web sites simultaneously for data collection, each targeting different segments of the study population. Kwok et al. (2009) proposed the use of multiple frame estimators for cross-population comparisons in a large data context.

Despite several advantages of MF surveys mentioned above, there are significant challenges in efficient estimation of population parameters of interest. In fact, the possible overlap among the frames at the selection stage leads to multiple opportunities of selection of the same unit in the final sample (although duplications may or may not occur in the sample) and hence the need to deal with multiplicity of selected units and their possible duplications at the estimation stage. In some special cases, the multiplicity problem can be overcome by removing the overlap among frames; i.e., de-duplicating the frames if it can be identified in advance at the design stage before the sample selection (Gonzales et al., 1996) or after the sample selection at the data collection stage by screening out the respondents if they had a chance of being selected in another sampling frame (Bankier, 1986). However, screening operations can be resource-consuming, error-prone, and essentially amount to missed opportunity to collect data from a willing participant. We therefore consider the general problem of MF estimation where each frame sample may contribute observations to applicable overlap domains, and although data from units are collected only once even if they are multiply selected, the data in the combined sample may not be de-duplicated. Research on MF estimation has been active since the seminal papers of Hartley, and several MF estimators have appeared in the literature, each derived under a somewhat different approach to estimation. This lack of a unified principled approach to MF estimation has not helped in making it popular among practitioners. Moreover, the natural complexity of the multiple frame estimation framework leads to difficulty in a simple interpretation and implementation of MF estimators including variance estimation without the availability of a standard software. This may be the main reason for limited use of MF designs despite its practical appeal.

The purpose of this paper is to provide a simplified and unified review of available MF estimation methods. This is accomplished by using the multiplicity approach to MF estimation as proposed in Singh and Mecatti (2009, 2011). The paper is organized as follows. In section 2, we present a simplified notation applicable to the general MF setup. Section 3 considers an appraisal of the concept of multiplicity to illustrate the relation between MF surveys and other techniques for dealing with imperfect frames such as Network Sampling and Indirect Sampling. In section 4, we illustrate how the multiplicity-approach offers a simple and direct way to generalize the familiar Horvitz-Thompson (HT) estimator to MF surveys. The class of HT-type MF estimators (termed GMHT, generalized multiplicity-adjusted HT estimators) is reviewed in section 5. It is observed that the GMHT class of Singh and Mecatti can encompass all unbiased or nearly unbiased MF estimators proposed in the literature under different approaches to MF estimation. Variance estimation is discussed in section 6. Finally, section 7 outlines future research directions with concluding remarks.

2. From Dual Frames to Multiple Frames: a simplified notation

For a good introductory account of the special case of two or dual frame (DF) surveys, see the review paper by Lohr (2009). The primary focus of the present paper is the general MF case. For a long time since Hartley first introduced the subject in the sixties, the literature has essentially offered estimation strategies for the DF case, usually alluding only in general terms to a possible extension to the MF case. It is speculated that the main reason for this is the lack of an easily understandable and generalizable notation for the MF setup as the traditional DF notation quickly becomes unwieldy for three or more frames as shown in Figure 1.

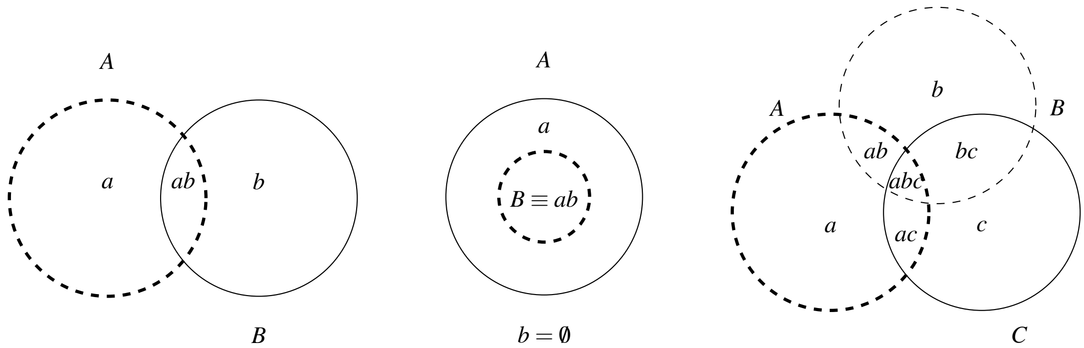

Figure 1: Two possible 2-frame survey cases and the general case of a 3-frame survey

The estimation theory for both DF and MF surveys has traditionally treated the problem of overlapping frames and possibly data duplication in the sample by partitioning the union of all frames into disjoint subsets (termed domains) as shown in Figure 1. The number of such domains increases exponentially with the number of frames involved in the survey. Figure 1 depicts the two possible scenarios of a DF survey as well as a general 3-frame case, using the traditional notation of capital letters to denote the frames and a combination of lowercase letters to denote the domains formed by intersecting frames. For instance, in the simplest DF case on the left side of Figure 1, domain a represents the set of units included in frame A only, domain b is the set of units included in frame B only and domain ab is the overlap domain, i.e. the set of doubly-counted units, included both in A and B .

A recent paper of Lohr and Rao (2006) is the first one to provide a general MF notation allowing for explicit formulae in a closed form for MF estimation based on domain-classifications. However their notation does not provide a straightforward generalization of the familiar DF notation as

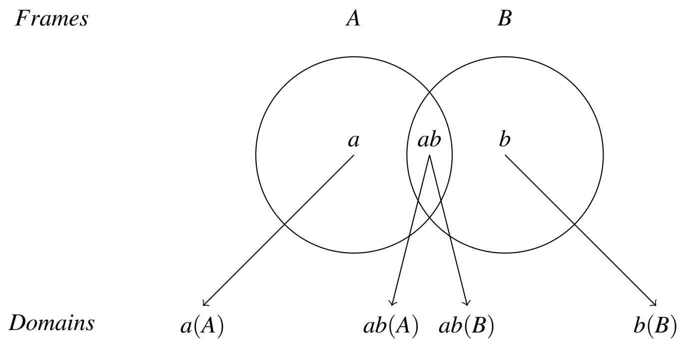

Figure 2: Partitioning of 2-Frame sample units into 4 frame-specific not necessarily disjoint domains

it relies on index sets leading to fairly complex matrix structure that can be difficult to interpret and implement. Moreover, as a consequence of the complex analytic form of the estimators, the analytical form of variance estimators also becomes quite complex.

We now propose a simplified notation applicable to the general MF estimation problem which is a natural generalization of the familiar DF notation.

In MF surveys, samples are generally selected independently from each frame. For every selected unit in each frame-specific sample, information about the membership in any of the other frames is typically collected besides data about the study variable(s). Assuming this information is both available and reliable, it allows for a classification of each frame-specific sample data into disjoint domains although domains may overlap across samples. This domain classification is a useful device for treating overlap domains without the need for de-duplication of sample data from common domains. In Figure 2 the domain-classification of sample data is sketched in the simplest DF case under the familiar DF notation: the two independent frame-specific samples, one from frame A and the other from frame B, generate four frame-specific domain samples, $a(A)$ and $ab(A)$ from frame A; $b(B)$ and $ab(B)$ from frame B.

Figure 3 illustrates a step-by-step generalization of this familiar DF notation first to a 3-frame setup (also compare with Figure 1) and then to an arbitrary number $Q \geq 2$ of frames.

As per Figure 3, we will use the following simplified notation. Let $U_1 \cdots U_q \cdots U_Q$ denote the collection of $Q \geq 2$ frames available for the survey. Assuming that the frame-membership information is correctly collected, the sample data from the frame U_q can be classified into a number D_q of disjoint domains $U_{1(q)} \cdots U_{d(q)} \cdots U_{D_q(q)}$. Notice that in the special case of all domains being non-empty, $D_q = 2^{Q-1}$ is constant for each frame $q = 1 \cdots Q$. For example for the 3-frame setup in the right side of Figure 1 , we have $D_q = 4$ non-empty domains for every frame $U_q,\ q = 1,2,3$ as also listed in the central panel of Figure 3.

It is seen that with the help of above notation along with the multiplicity approach as discussed in the next section, it is possible to develop a simplified and unified MF estimation theory encompassing the main methods in the literature (Sections 4 and 5) as proposed by Singh and Mecatti (2009, 2011). This simplification also allows for a closed form HT-type expression of variance estimators as shown in Section 6.

	2 frames (DF)	
frames →	A	B
domains →	a(A)	b(B)
	ab(A)	ab(B)

3-frames		
A	B	C
a(A)	b(B)	c(C)
ab(A)	ab(B)	ac(C)
ac(A)	bc(B)	bc(C)
abc(A)	abc(B)	abc(C)

Q-frames (MF)	
U_q	$q = 1 \ldots Q$
$U_{1(q)}$	
$\vdots$	
$U_{d(q)}$	$d = 1 \ldots D_q$
$\vdots$	
$U_{D_q(q)}$	

Figure 3: Partitioning of multi-frame samples into frame-specific domains

3. Multiplicity as a unifying estimation strategy for Network sampling, Indirect sampling, and Multiple Frame surveys

The concept of multiplicity was first used by Birnbaum and Sirken (1965) in introducing the Network sampling as an estimation strategy for surveying rare or elusive populations. Network sampling, also known as Multiplicity or Snowball sampling, is useful, for instance, in estimating the prevalence of a rare disease where a single frame representing the target population is typically not available. Instead, a household list that may be easily available, can be used for selecting the sample. For every selected household, all the occupants are interviewed and also asked to report about other individuals related to them in some manner under a specified linkage rule such as the sibling relationship. In Figure 4, a sketch of the sampling situation described above is given. Households present in the list or the sampling frame (different from the target frame) act as a selection device and are linked to the target units by the specified linkage rule. Selection units and target units may have no relation, or may be related according to a linkage pattern which can be one-to-one (units could be identical in particular) or one-to-many or many-to-one. When a household is selected, all units in the network of linked target units, if any, are eligible for data collection and included in the final sample. Consequently, target units linked with more than one household, such as individuals with several siblings living in separate households, are in a sense overcounted at the population level and therefore have a higher probability of being multiply included in the final sample than individuals with few or no siblings. Multiplicity is defined for every target unit as the number of selection units to which it is linked to, and then a multiplicity-adjusted estimator can be defined. Generalizations to many-to-many linkage patterns can also be made as, for instance, the case of multiple-listing of the same unit into the sampling frame. See Sirken (2004) for an historical review of network sampling.

A more recent strategy termed Indirect sampling was introduced by Lavallée (2007) for dealing with similar imperfect sampling frame situations in the context of social and economic surveys. Due to lack of a sampling frame representing directly all the target population units, an indirect sampling frame is used instead. A typical example occurs in implementing surveys on small children when only a list of parent names is available as an indirect sampling frame. Since parents may live in separate houses, a setup similar to Figure 4 emerges. The generalized weight

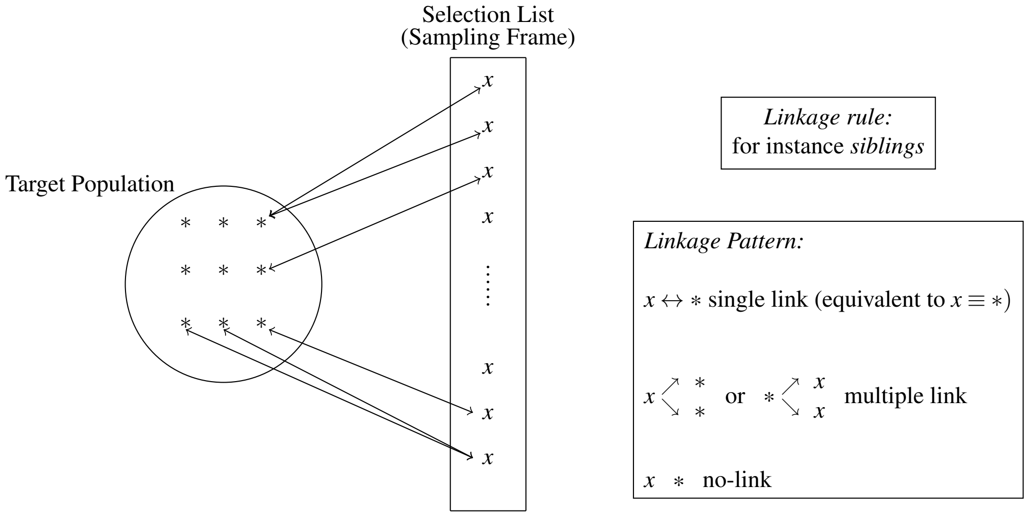

Figure 4: Target Population and Sampling Frame setup for both network sampling and indirect sampling

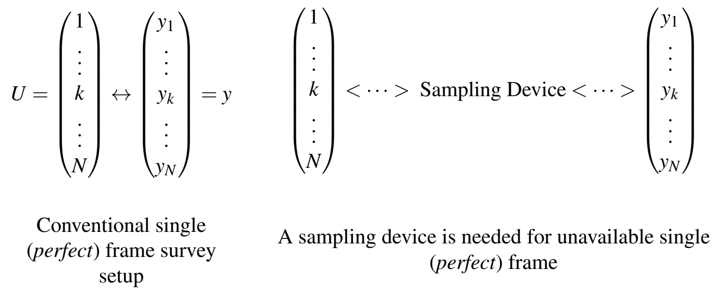

Figure 5: From *perfect* to *imperfect* frame survey setup

share method is suggested to adjust for the conceptual multiple inclusions of target units at the estimation stage (although the sample may have very few or no duplications) by providing a multiplicity-adjusted sampling weight for every selected unit in the sample. Other forerunners of the generalized weight share method are the variable weighting by Rao (1968) to handle a single frame with an unknown amount of duplication, and weight-sharing by new selected units linked longitudinally with units selected in earlier waves as introduced by Ernst (1989).

We now observe that both network and indirect sampling frameworks essentially refer to the same conceptual layout exemplified in Figure 4 which can also be deemed to be applicable to the MF framework. This is explained in Figure 5. The left side of Figure 5 illustrates the conventional sampling framework based on a conceptual perfect single-frame of labelled population units. The collection of labels $(k = 1 \cdots N)$ then identifies the target population U of size N and at the same time defines a perfect one-to-one linkage rule between each target unit k and the observable value y_k of the study variable. In an imperfect frame situation, the one-to-one linkage rule is no longer present so that a sampling device is needed for selection purposes, as displayed in the right side

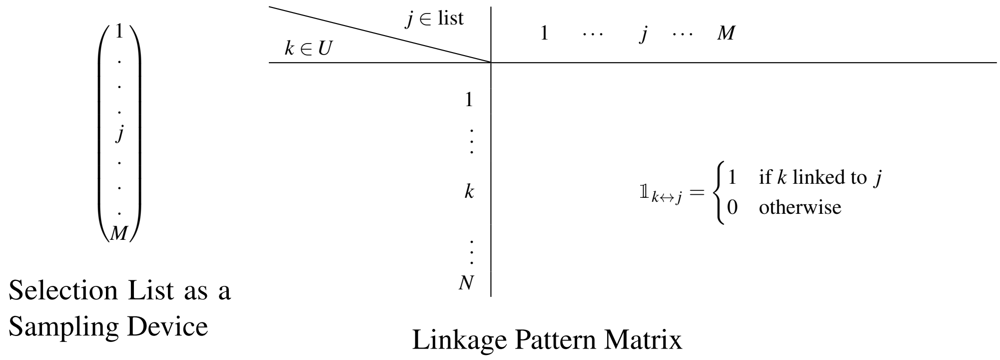

Figure 6a: Sampling Device and Linkage Pattern Matrix in both Network and Indirect sampling

of Figure 5. The sampling device may be a collection of selection units linked in a well-defined manner to target units as in network sampling or as in indirect sampling. In both these cases, the sampling device identifies target units with a set of selection units $(j = 1 \cdots M)$. When the sampling device is a collection of $Q \geq 2$ sampling frames, possibly incomplete and overlapping as shown in Figure 1, we have an MF sampling framework.

The linkage pattern between the target population and the sampling device drives the actual data collected for target units. The nature of linkage may vary with the sampling design: it is generally one-to-many in network sampling; it can be many-to-many for indirect sampling in order to allow some sort of clustering within the sampling device; and it is one-to-one within each frame and many-to-one between frames in MF surveys. Considering first both network and indirect sampling designs, regardless of the complexity of the linkage pattern, it is completely described by an $N \times M$ linkage matrix whose entries are non-random indicators $\mathbb{1}_{k \leftrightarrow j}$ taking values of 1 if the population unit k is linked to the selection unit j and 0 otherwise (see Figure 6a).

Now consider the multiplicity approach for estimation for network and indirect sampling designs. Multiplicity, as first introduced by Birnbaum and Sirken (1965), is defined for every population unit as the sum over the rows of the linkage matrix in Figure 6a, namely $m_k = \sum_{j=1}^{M} \mathbb{1}_{k \leftrightarrow j}$. Notice that for a conventional single frame survey, we would set $m_k = 1$ for every $k \in U$. In network sampling $m_k > 1$ for at least one unit $k \in U$ and $\mathbb{1}_{k \leftrightarrow j}$ is termed as the multiplicity counting rule. That is why network sampling is also known as multiplicity sampling.

The frame-specific linkage rule involved in an MF survey is the frame membership indicator $\mathbb{1}_{k \in U_q}, q = 1 \cdots Q$, taking values of 1 if population unit k is included in frame U_q and 0 otherwise, leading to the frame-linkage matrix in Figure 6b where rows represent the unavailable single frame units and columns display the $Q \geq 2$ available frames.

In an MF context unit multiplicity, i.e. the sum over each row, is then the number of frames in which every unit belongs to

$$m_k = \sum_{q=1}^{Q} \mathbb{1}_{k \in U_q} \tag{1}$$

while the sum over the columns gives the frame size $N_q = \sum_{k=1}^{N} \mathbb{1}_{k \in U_q}$. It follows that $\sum_{q=1}^{Q} N_q \geq N$ in the general case of overlapping frames.

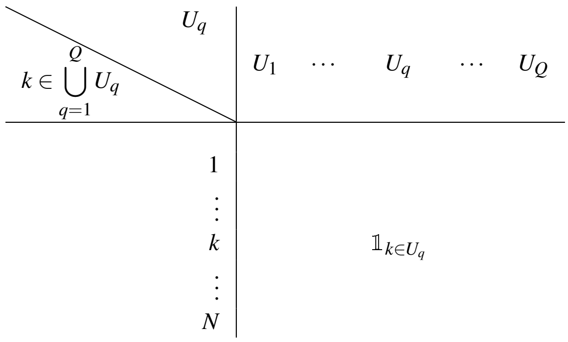

Figure 6b: Frame membership indicator as linkage rule for a MF survey

Finally, disjoint domains are formed by counting only once all identical rows; i.e. the collection of population units sharing the very same array of frame-membership indicators (compare Figures 3 and 6b). Consequently, all units included in the same domain share the same multiplicity, thus multiplicity can be defined either as a unit characteristic, as in definition (1), or also as a domain characteristic; i. e., the number of frames intersecting with a given domain. In Figure 7, a simple example is illustrated for a population of size $N = 10$ consisting of 3 overlapping frames. As mentioned in Section 1, domains play a major role in the development of the MF estimation theory in that the domain classification of sample data is used to handle the overlapping frame issue arising in estimation with MF surveys. The target population and sampling frame linkage setup described above, besides providing a unified framework of three scenarios for dealing with imperfect frames (MF, Network and Indirect sampling), allows for a multiplicity approach to all three scenarios.

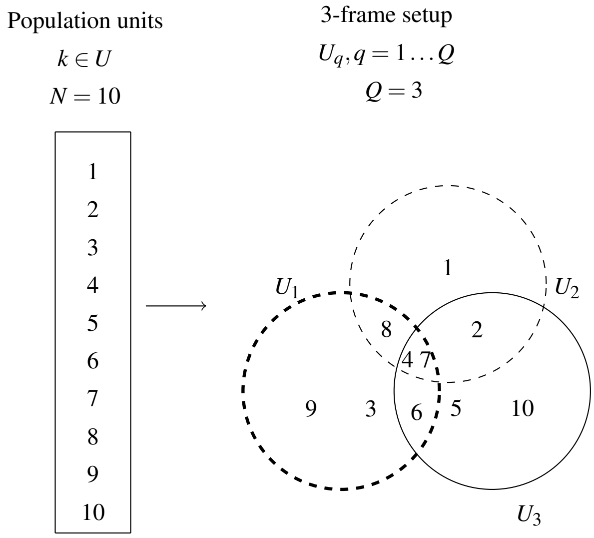

Figure 7 : Example of the domain classification in a 3-frame setup (1/2)

Linkage Matrix $\mathbb{1}_{k \in U_q}$

k \ U_q	U_1	U_2	U_3	m_k
1	0	1	0	1
2	0	1	1	2
3	1	0	0	1
4	1	1	1	3
5	0	0	1	1
6	1	0	1	2
7	1	1	1	3
8	1	1	0	2
9	1	0	0	1
10	0	0	1	1
Frame size N_q	6	5	6	

→

$2^Q - 1 = 7$ frame membership arrays	Population units sharing the same array (Domain d)	Domain multiplicity m_d
100	{3,9}	1
010	{1}	1
001	{5,10}	1
110	{8}	2
101	{6}	2
011	{2}	2
111	{4,7}	3

→

Frame-specific Domain classification:
each frame U_q consists of $D_q = 4$ domains $U_{d(q)}$
$d = 1 \dots D_q$

$U_{d(q)}$ \ U_q	U_1	U_2	U_3
→	$U_{1(1)} = \{3,9\}$	$U_{2(1)} = U_{1(2)}$	$U_{1(3)} = U_{3(1)}$
	$U_{2(1)} = \{8\}$	$U_{2(2)} = \{1\}$	$U_{2(3)} = U_{3(2)}$
	$U_{3(1)} = \{6\}$	$U_{3(2)} = \{2\}$	$U_{3(3)} = \{5,10\}$
	$U_{4(1)} = \{4,7\}$	$U_{4(2)} = U_{4(1)}$	$U_{4(3)} = U_{4(2)} = U_{4(1)}$

Figure 7 : Example of the domain classification in a 3-frame setup (2/2)

4. Multiplicity Approach for Generalizing HT Estimation to Multiple Frame Surveys

In an MF survey, Q frame-samples $\{s_q\}$ are independently selected under possibly different designs. At the estimation stage, the collection of data from the Q samples is used to produce an estimate of the population parameter where the target population is represented by the union of all frames. As a first step toward a simplified and unified approach to MF estimation, we consider a broad classification of MF estimators available in the literature according to two basic approaches. One approach relies on first concatenating samples from all frames into a single combined sample with suitable weights and then to directly compute estimates of population parameters. In the other approach, first separate estimates of each domain are computed by using each sample $s_{d(q)}$ falling in the domain $U_{d(q)}$. Domain estimates are then aggregated over all the domains within and between frames (i.e., $\sum_q \sum_d$)) to obtain an estimate of the population parameter. This is somewhat analogous to a stratified estimator by treating the domain classification as a virtual partition of the target population in each frame except that domain estimates from the same frame are not independent.

The two approaches will be referred to respectively as *combined* frame approach (COMB) and *separate* frame approach (SEP) as in Singh and Wu (2003), Singh and Mecatti (2011) and

Singh and Mecatti (2014). A COMB estimator is also sometimes referred to as a single frame estimator (see for instance Lohr, 2009). However, we prefer the term combined frame (COMB) to avoid ambiguity with the traditional single frame estimators including a special single frame estimation approach to MF in which duplicate selected units in the combined sample are first eliminated, and then for all sampled units from overlapping frames, revised inclusion probabilities are computed to adjust for the possibility of being selected from several frames as input to HT estimation (Bankier, 1986).

In Singh and Mecatti (2011), an additional practical classification is proposed, based on the amount of frame-level information available for each sampled unit at the estimation stage, termed basic, partial, or full. The frame-level information is *in addition* to the usual sample data collected from respondents. Specifically, basic frame level information refers to the case if for every sampled unit $k \in s_q$ selected from a given frame U_q

1. the unit multiplicity m_k is collected while the frame identification $\mathbb{1}_{k \in U_{q'}}$ may not be available for all the other frames $q' \neq q = 1 \cdots Q$;
2. the inclusion probability is known only for selection in U_q while it may not be available for all the other frames $q' \neq q$.

The case of partial frame level information refers to the situation when we have basic information plus

3. identification of all frame memberships from all the frames the sampled unit could have come from, i.e. $\mathbb{1}_{k \in U_q}$ is given for all $q = 1 \cdots Q$.

Finally, the case of full frame level information refers to the situation when we have partial information plus

4. the inclusion probabilities for every sampled unit $k \in s_q$ are available for all frames $q = 1 \cdots Q$, namely *in addition* to U_q in which k was actually sampled.

In other words the basic frame level information is essentially restricted to the frame in which the unit was actually sampled while very limited information is used from other frames which might include the same unit. This happens, for instance, when surveying sensitive characteristics (such as personal habits) due to confidentiality concerns. It might also arise when using stigmatized frames such as lists of mental patients or of convicted people. Units might be sensitive to disclosing information about the other frame membership and refuse to answer when asked about "*which* other frame they belong to" besides the one in which they are being interviewed. In these cases, unit multiplicity, which refers to the number of frames a unit belongs, might be easier to collect using a more discreet "*how many* frames" question. Notice that the basic information does not allow for classifying sample data into domain-samples as exemplified in Figure 7, and therefore all the available MF estimators become inapplicable except for the simple multiplicity-adjusted estimator (Mecatti, 2007). By contrast, both partial and full frame level information imply having more information about all the Q frames involved in the survey, and both allow classifying the frame-specific sample data into domain-samples.

A SEP estimator requires basic or partial frame level information (i.e., not full information) for each unit. Most of the available MF estimators are in the SEP class. In particular, members of this class include the optimal DF Hartley estimator (Hartley, 1962, 1974) and its improvements for simple random sampling designs using maximum likelihood arguments by Lund (1968), Fuller and Burmeister (1972) and Skinner (1991); DF pseudo maximum likelihood estimator (PML) of Skinner and Rao (1996), generalizations to MF by Lohr and Rao (2006) and the simple multiplicity estimator of Mecatti (2007); and modified regression estimators of Singh and Wu (1996, 2003). On the other hand, a COMB estimator requires full frame level information and this class encompasses methods by Kalton and Anderson (1986) and Bankier (1986).

From the forms of SEP and COMB estimators, it was observed by Singh and Mecatti (2009, 2011) that for MF estimators having the form of an expansion or calibration estimator, a unified and simplified approach to MF estimation for the problem of combining data from the Q independent frame-samples s_q can be developed by using a weighting system adjusted for multiplicity. This will be explained in the context of estimating total parameters, starting below with the generalization to MF of the familiar single frame HT estimator by using unit multiplicity (1), and then proceeding in Section 5 to illustrate both SEP and COMB estimators as generalized multiplicity-adjusted HT estimators.

In a conventional single-frame survey, for estimating the population total $Y = \sum_{k\in U} y_k$, the familiar Horvitz-Thompson estimator is computed by using data from a single sample s as

$$\hat{Y}_{HT} = \sum_{k\in s} y_k \pi_k^{-1} \tag{2}$$

with initial design weights π_k^{-1} obtained as inverse of the (first order) inclusion probabilities. In an MF survey the target population is covered by $Q \geq 2$ frames so that the population total is given by

$$Y = \sum_{k\in \bigcup_{q=1}^{Q} U_q} y_k \tag{3}$$

where .

In the MF case, when frames can be incomplete and overlapping, the total parameter using unit multiplicity can be expressed as

$$Y = \sum_{q=1}^{Q} \sum_{k\in U_q} y_k m_k^{-1} \tag{4}$$

A simple design unbiased multiplicity estimator for (4) follows in a straightforward manner and is given by (see Mecatti, 2007; Lohr, 2011)

$$\hat{Y}_{SM} = \sum_{q=1}^{Q} \sum_{k\in s_q} y_k m_k^{-1} \pi_k^{-1} \tag{5}$$

Expression (5) shows how the (inverse of) unit multiplicity is used to adjust the design weight to avoid bias due to the possibility of inclusion from more than one frame and of possible duplication in an MF survey. In a conventional single frame survey, $m_k = 1$ for all $k = 1 \cdots N$ and equation (5) reduces to the HT estimator (2). The case of Q non-overlapping frames is equivalent to a target population partitioned into Q strata. In this case, every population unit has the multiplicity of 1 and equation (5) reduces to the usual stratified HT estimator.

Unit multiplicity is a natural choice for adjusting HT estimation to MF surveys. However, it is not the only way to obtain unbiased estimates. By defining a general multiplicity-adjustment factor, a generalized multiplicity-adjusted HT class of estimators for MF surveys is presented in the next section.

5. The generalized multiplicity-adjusted HT class as a simplified and unified approach to multiple frame estimation

In Singh and Mecatti (2009, 2011) a generalized HT methodology for MF estimation is given under a general definition of multiplicity adjustment. Let $\alpha_{k(q)}$ be the multiplicity adjustment factor for every unit k in a given frame U_q with $\sum_q \alpha_{k(q)} = 1$. Generally $\alpha_{k(q)} \in [0,1]$ although it is not necessary. From equation (5), a *Generalized Multiplicity-adjusted Horvitz-Thompson* (GMHT) class of MF estimators is defined as

$$\hat{Y}_{GMHT} = \sum_{q=1}^{Q} \sum_{k \in s_q} y_k \alpha_{k(q)} \pi_{k(q)}^{-1} \tag{6}$$

where the factor $\alpha_{k(q)}$ ensures that y_k is counted only once even if unit k is present in more than one frame. Notice that a GMHT estimator is design-unbiased by construction. Moreover equation (6) shows that any GMHT estimator is a linear combination of independent HT-type estimators; therefore conventional HT-type variance estimator can be easily obtained in a closed form as shown in the next section.

Different GMHT estimators are derived by different choices of the multiplicity adjustment α-factor in (6). The multiplicity-adjusted MF estimator as given in (5) is the simplest GMHT estimator with the basic choice of $\alpha_{k(q)} = m_k^{-1}$; i.e., a constant adjustment regardless of the frame U_q that contains the unit k. The GMHT class includes all the known design-unbiased or approximately unbiased MF estimators, either COMB or SEP, whether requiring basic, partial, or full information according to the classification illustrated in the previous section. Singh and Mecatti (2009, 2011) introduced the GMHT class to deal with the case of mixed frame-level information when some units had only basic while others had full information, and also considered variations of the COMB approach. The COMB approach involves unit/frame-specific multiplicity adjustments, or α-factors. For instance, it can be easily shown that the Kalton-Anderson estimator (Kalton and Anderson, 1986) is a GMHT estimator. Consider, for simplicity, a DF survey with equal probability selection such as simple random sampling from both frames. Kalton and Anderson proposed to weight data from the overlap domain by inclusion probabilities from both frames as follows

$$\hat{Y}_{KA} = \hat{Y}_{a(A)} + \frac{\pi_A}{\pi_A + \pi_B} \hat{Y}_{ab(A)} + \frac{\pi_B}{\pi_A + \pi_B} \hat{Y}_{ab(B)} + \hat{Y}_{b(B)} \tag{7}$$

where, under equal probability designs, $\pi_{k(A)} = \pi_A$ and $\pi_{k(B)} = \pi_B$

$\hat{Y}_{a(A)} = \sum_{k \in s_A} y_k \pi_A^{-1} \mathbb{1}_{k \in a}$ is the HT estimator of the total of domain a;

$\hat{Y}_{ab(A)} = \sum_{k \in s_A} y_k \pi_A^{-1} \mathbb{1}_{k \in ab}$ and $\hat{Y}_{ab(B)} = \sum_{k \in s_B} y_k \pi_B^{-1} \mathbb{1}_{k \in ab}$ are both HT estimators of the same overlap domain total; and

$\hat{Y}_{b(B)}$ is defined accordingly (also refer to Figures 1 and 2).

Thus the Kalton-Anderson estimator adjusts data from the overlap domain by giving higher weight to the estimator from the frame with higher inclusion probability. Notice that for non-overlap domains a and b, the unit multiplicity adjustment $m_k = 1$ is used.

With little algebra, equation (7) can be re-expressed as a sum of contributions from the two frame-samples as give below

$$\hat{Y}_{KA} = \sum_{k \in s_A} y_k \pi_A^{-1} \left[\mathbb{1}_{k \in a} + \pi_A (\pi_A + \pi_B)^{-1} \mathbb{1}_{k \in ab} \right] + \sum_{k \in s_B} y_k \pi_B^{-1} \left[\mathbb{1}_{k \in b} + \pi_B (\pi_A + \pi_B)^{-1} \mathbb{1}_{k \in ab} \right] \tag{8}$$

which can easily be generalized to MF with $Q \geq 2$ frames under a general sampling design in each frame; i.e., for unequal probability designs with inclusion probabilities $\pi_{k(A)}$ and $\pi_{k(B)}$. Thus, using the simplified notation introduced in Section 2, equation (8) generalizes to (refer to Figure 3)

$$\hat{Y}_{KA} = \sum_{q=1}^{Q} \sum_{k \in s_q} y_k \pi_{k(q)}^{-1} \left\{ \sum_{d=1}^{D_q} \pi_{k(q)} \left[\mathbb{1}_{k \in U_{d(q)}} \left(\sum_{q'=1}^{Q} \pi_{k(q')} \mathbb{1}_{k \in U_{d(q')}} \right)^{-1} \right] \right\} \tag{9}$$

Both equations (8) and (9) show that for the estimator $\hat{Y}_{KA}$, the following needs to be computed

i. a complete identification of frame membership for every sampled units k in order to compute $\mathbb{1}_{k \in U_{d(q)}}$ and $\mathbb{1}_{k \in U_{d(q')}}$ which track across frames the domains that include unit k. and

ii. a complete knowledge of inclusion probability for every frame in which unit k could have been selected in order to compute the sum across frames $\sum_{q'=1}^{Q} \pi_{k(q')} \mathbb{1}_{k \in U_{d(q')}}$.

It follows that $\hat{Y}_{KA}$ is applicable if full frame-level information is available as defined in Section 4. Also it is readily seen that $\hat{Y}_{KA}$ is a GMHT estimator as defined in (6) with the following choice of the multiplicity-adjustment α-factor

$$\alpha_{k(q)}^{KA} = \pi_{k(q)} \sum_{d=1}^{D_q} \left[\mathbb{1}_{k \in U_{d(q)}} \left(\sum_{q'=1}^{Q} \pi_{k(q')} \mathbb{1}_{k \in U_{d(q')}} \right)^{-1} \right] \tag{10}$$

Notice that $\alpha_{k(q)}^{KA} = 1$ for unit k with multiplicity $m_k = 1$; i.e., for unit included in one and only one frame U_q. Furthermore $\sum_{q=1}^{Q} \alpha_{k(q)}^{KA} = 1$ for every population unit k.

GMHT estimation also applies to the SEP approach to MF estimation. A SEP estimator requires partial frame level information, and its computation basically consists of the following steps:

i. Consider separately data from each frame-sample s_q after classifying into domain-samples $s_{d(q)}, d = 1 \cdots D_q$;

ii. Compute estimates of domain totals and combine multiple estimates as the case may be; and

iii. Aggregate all domain estimates to produce an overall estimate of the population total.

It is observed that a domain-specific (and not unit-specific) multiplicity-adjustment is needed in a SEP estimator. The SEP estimator belongs to the GMHT class (6) with multiplicity-adjustment α-factor given by

$$\alpha_{k(q)} = \sum_{d=1}^{D_q} \alpha_{d(q)} \mathbb{1}_{k \in U_{d(q)}} \tag{11}$$

which is common for all units in the same domain $U_{d(q)}$ such that for every population unit k

$$\sum_{q=1}^{Q} \sum_{d=1}^{D_q} \alpha_{d(q)} \mathbb{1}_{k \in U_{d(q)}} = \sum_{q=1}^{Q} \alpha_{k(q)} = 1 \tag{12}$$

Expression (11) is applicable, for instance, to the Hartley's optimal unbiased estimator (Hartley, 1962, 1974) for DF under general designs for both frames. That is,

$$\hat{Y}_H = \hat{Y}_{a(A)} + \alpha \hat{Y}_{ab(A)} + (1-\alpha)\hat{Y}_{ab(B)} + \hat{Y}_{b(B)} \tag{13}$$

with α chosen to minimize the variance $V\left(\hat{Y}_H\right)$ and is given by

$$\alpha^H = \frac{V\left(\hat{Y}_{ab(B)}\right) + Cov\left(\hat{Y}_{b(B)}, \hat{Y}_{ab(B)}\right) - Cov\left(\hat{Y}_{a(A)}, \hat{Y}_{ab(A)}\right)}{V\left(\hat{Y}_{ab(A)}\right) + V\left(\hat{Y}_{ab(B)}\right)} \tag{14}$$

Estimator (13) can be reexpressed as a GMHT estimator analogous to the Kalton-Anderson estimator (compare equations (7) and (8))

$$\begin{aligned} \hat{Y}_H = & \sum_{k \in s_A} y_k \pi_A^{-1} \left(\mathbb{1}_{k \in a} + \alpha^H \mathbb{1}_{k \in ab}\right) \\ & + \sum_{k \in s_B} y_k \pi_B^{-1} \left(\mathbb{1}_{k \in b} + (1-\alpha^H) \mathbb{1}_{k \in ab}\right) \end{aligned} \tag{15}$$

and can be generalized to $Q \geq 2$ frames as

$$\hat{Y}_H = \sum_{q=1}^{Q} \left[\sum_{k \in s_q} y_k \pi_{k(q)}^{-1} \left(\sum_{d=1}^{D_q} \alpha_{d(q)}^H \mathbb{1}_{k \in U_{d(q)}} \right) \right] = \sum_{q=1}^{Q} \sum_{k \in s_q} y_k \pi_{k(q)}^{-1} \alpha_{k(q)}^H \tag{16}$$

Thus Hartley's MF estimator is GMHT, with multiplicity-adjustment α-factor given by (11) where optimal $\alpha_{d(q)}^H$ minimizes $V\left(\hat{Y}_H\right)$ under the constraint (12). In practice, it is only approximately unbiased and optimal because $\alpha_{d(q)}^H$ is estimated from the sample; see also simulation results (Lohr and Rao, 2006; Mecatti, 2007). Using regression on zero-functions (Fuller and Burmeister, 1972; Singh and Wu, 1996), a simpler expression for the optimal $\alpha_{d(q)}^H$ for MF can be obtained. First we recall, from the comment following equation (1), that each and every unit from the same domain shares the same multiplicity, so that the unit multiplicity m_k, for all $k \in U_{d(q)}$, also defines the multiplicity of the domain $U_{d(q)} \ni k$. Moreover, domain multiplicity defines a domain characteristic regardless of *which* frame U_q contains that domain. When referred to domain, multiplicity equals *how many* frames intersect with that domain and can be denoted as $m_d, \quad d = 1 \cdots D,$ where $D \leq 2^Q - 1$ is the number of non-empty domains generated by the Q frames used in the survey.

Second we notice that m_d estimators $\hat{Y}_{d(q)}$ are available for the same domain total, one from each frame-sample s_q intersecting with the domain.

Finally using the equivalence between domain multiplicity m_d and unit multiplicity m_k common to every unit k included in that domain, we observe that the simple multiplicity-adjusted estimator, based on m_k as defined in (5), can also be re-expressed as a linear combination of HT domain total estimators adjusted by domain multiplicity

$$\hat{Y}_{SM} = \sum_{d=1}^{D} \sum_{q=1}^{m_d} m_d^{-1} \hat{Y}^{HT}_{d(q)} \tag{17}$$

It follows that the Hartley's (SEP) estimator, as defined in (16), can also be obtained by combining the m_d independent HT estimators available for the same domain total except for domains with $m_d = 1$. It can be expressed as a regression estimator, by regressing $\hat{Y}_{SM}$ on predictor zero functions formed by pairs of estimators for the same domain from different frames (see Fuller and Burmeister, 1972, Singh and Mecatti, 2014 and Singh and Wu, 2003)

$$\hat{Y}_H = \hat{Y}_{SM} - \sum_{d=1}^{D} \left[\sum_{q=1}^{m_d} \sum_{q'>q=1}^{m_d} \beta_{d(qq')} \left(\hat{Y}^{HT}_{d(q)} - \hat{Y}^{HT}_{d(q')} \right) \right] \tag{18}$$

where $\hat{Y}^{HT}_{d(q)} = \sum_{k \in s_{d(q)}} y_k \pi^{-1}_{k(q)}$ is the HT estimator of the total of domain $U_{d(q)}$ computed with data from frame-sample s_q classified into domain-samples $s_{d(q)} = s_q \bigcap U_{d(q)}, \ d = 1 \cdots D_q$.

The optimal solution for the regression coefficient $\beta_{d(qq')}$ in equation (18) then follows from optimal regression of $\hat{Y}_{SM}$ on zero functions and can be readily computed by standard software for linear regression analysis. Let $\varphi_{d(qq')} = \hat{Y}^{HT}_{d(q)} - \hat{Y}^{HT}_{d(q')}$. Notice that $m_d (m_d - 1)/2 = \binom{m_d}{2}$ predictors $\varphi_{d(qq')}$ are available so that a predictor vector $\underline{\varphi}$ of dimension $H = \sum_{d=1}^{D} m_d (m_d - 1)/2 = \sum_{d=1}^{D} \binom{m_d}{2}$ can be defined. Let $V_{\varphi\varphi}$ define the Var-Cov matrix of $\underline{\varphi}$ and $Cov\left(\underline{\varphi}, \hat{Y}_{SM}\right)$ be the $(H \times 1)$ vector of covariances between every predictor in $\underline{\varphi}$ and $\hat{Y}_{SM}$. We have

$$Cov\left(\varphi_{d(qq')}, \hat{Y}_{SM}\right) = \sum_{d'=1}^{D} \sum_{q''=1}^{m_{d'}} m_{d'}^{-1} \left[Cov\left(\hat{Y}^{HT}_{d(q)}, \hat{Y}^{HT}_{d'(q'')}\right) - Cov\left(\hat{Y}^{HT}_{d(q')}, \hat{Y}^{HT}_{d'(q'')}\right) \right] \tag{19}$$

for all $q'' = 1 \cdots m_d, \ q' \neq q$, and the (h, h') element of $V_{\varphi\varphi}$, for $h \neq h' = 1 \cdots H, \ h = d(qq'), \ h' = d(q''q''')$, is given by

$$Cov(\varphi_h, \varphi_{h'}) = Cov\left(\hat{Y}^{HT}_{d(q)} - \hat{Y}^{HT}_{d(q')}, \hat{Y}^{HT}_{d(q'')} - \hat{Y}^{HT}_{d(q''')}\right)$$

$$= Cov\left(\hat{Y}^{HT}_{d(q)}, \hat{Y}^{HT}_{d(q'')}\right) - Cov\left(\hat{Y}^{HT}_{d(q)}, \hat{Y}^{HT}_{d(q''')}\right) - Cov\left(\hat{Y}^{HT}_{d(q')}, \hat{Y}^{HT}_{d(q'')}\right) + Cov\left(\hat{Y}^{HT}_{d(q')}, \hat{Y}^{HT}_{d(q''')}\right). \tag{20}$$

Similarly, covariance terms for all the other scenarios such as

$$h = d(qq'),\ h' = d(qq'');\quad h = d(qq'),\ h' = d'(qq');\quad \text{and}\ \ h = d(qq'),\ h' = d'(q''q''')$$

can be obtained. Thus an explicit closed-form expression for the $(H \times 1)$ vector $\underline{\beta}$ of optimal regression coefficients $\beta_{d(qq')}$ in (18) is given by

$$\underline{\beta} = V_{\varphi\varphi}^{-1}\, Cov\left(\underline{\varphi}, \hat{Y}_{SM}\right) \tag{21}$$

Now observe that the estimator (18) can also be expressed as

$$\hat{Y}_H = \sum_{d=1}^{D}\sum_{q=1}^{m_d}\left(m_d^{-1} - \sum_{q'\neq q=1}^{m_d}\beta_{d(qq')}\right)\hat{Y}_{d(q)}^{HT} = \sum_{q=1}^{Q}\sum_{d=1}^{D_q}\left(m_d^{-1} - \sum_{q'\neq q=1}^{m_d}\beta_{d(qq')}\right)\hat{Y}_{d(q)}^{HT} \tag{22}$$

where $\beta_{d(qq')} = -\beta_{d(q'q)}$, so that

$$\alpha_{d(q)}^{H} = \begin{cases} 1 & \text{if}\ \ m_d = 1 \\ m_d^{-1} - \sum_{q'\neq q=1}^{m_d}\beta_{d(qq')} & \text{if}\ \ m_d \geq 2 \end{cases} \tag{23}$$

Notice that equation (23) generalizes the DF solution given in (14) and the unit sum constraint (12) is satisfied because, for every d, $\sum_{q=1}^{m_d} m_d^{-1} = 1$ and $\sum_{q=1}^{m_d}\sum_{q'\neq q=1}^{m_d}\beta_{d(qq')} = 0$. For SEP estimators involving more general regression, see the comment in Section 7

6. Multiplicity-adjusted variance estimation of Multiple Frame estimators

Unbiased or approximately unbiased multiplicity-adjusted estimators included in the GMHT class (6) are in fact linear combinations of Q independent HT estimators - one from each frame-sample s_q. Thus the estimator's variance as well as its unbiased estimator follow directly from HT-type estimation as explained in Singh and Mecatti (2011). In other words, the simplification induced by adopting the multipicity approach, besides allowing for generalizing the customary HT estimation to MF surveys, also allows for standard variance estimation in a closed analytical form, which may involve linearization. The exact variance of any GMHT estimator in the well known HT form is given by

$$V\left(\hat{Y}_{GMHT}\right) = \sum_{q=1}^{Q}\left[\sum_{k\in U_q} y_k^2\alpha_{k(q)}^2\frac{1-\pi_{k(q)}}{\pi_{k(q)}} + \sum_{k\neq k'}\sum_{\in U_q}\frac{y_k\alpha_{k(q)}\, y_{k'}\alpha_{k'(q)}}{\pi_{k(q)}\pi_{k'(q)}}\left(\pi_{kk'(q)} - \pi_{k(q)}\pi_{k'(q)}\right)\right] \tag{24}$$

and the Sen-Yates-Grundy (SYG) form of variance estimator for fixed sample size designs is given by

$$v\left(\hat{Y}_{GMHT}\right) = \frac{1}{2}\sum_{q=1}^{Q}\sum_{k\neq k'}\sum_{\in s_q}\frac{\pi_{k(q)}\pi_{k'(q)} - \pi_{kk'(q)}}{\pi_{kk'(q)}}\left(\frac{y_k\alpha_{k(q)}}{\pi_{k(q)}} - \frac{y_{k'}\alpha_{k'(q)}}{\pi_{k'(q)}}\right)^2 \tag{25}$$

For example, for Hartley (optimal) estimator $\hat{Y}_H$, we first write the GMHT form (16), and then use (25) to obtain its variance estimator analogous to usual HT estimators. Resampling methods for variance estimation can also be easily applied by considering the variance of the contributions from each frame separately as given by the GMHT form (6).

7. Concluding remarks and future research directions

In this paper we considered the connection between MF sampling, Indirect sampling and Network sampling and showed how all estimators can be expressed as a multiplicity-adjusted estimator. MF estimators can be classified into two class, SEP and COMB. It was shown that the unbiased MF estimators - such as the Kalton-Anderson COMB estimator (9) and the Hartley SEP estimator with known α^H (16) - can be expressed in a GMHT form introduced earlier by Singh and Mecatti (2009, 2011). The GMHT class can be extended by relaxing the assumption of unbiasedness to approximate unbiasedness. For example, as mentioned earlier for Hartley estimator, it is only approximately umbiased and optimal because in practice α^H is estimated from the same data. Similarly, for regression calibration estimators, the weight $w_{k(q)} \left(= \pi_{k(q)}^{-1}\right)$ is adjusted by the factor $a_{k(q)}$ where $a_{k(q)} = 1 + O_p(1/\sqrt{n})$. By allowing $w_{k(q)}a_{k(q)}$ instead of $\pi_{k(q)}^{-1}$ in the GMHT class, the resulting estimators will be approximately unbiased in general. However, this way the new GMHT-reg class can subsume various regression-based MF estimators in the SEP class such as Fuller-Burmeister optimal regression estimator (Fuller and Burmeister, 1972 — whenever it can be expressed as a calibration estimator), pseudo-maximum estimator (PML) of Skinner and Rao (1996) and Lohr and Rao (2006) under a working covariance for regression after Hájek-ratio adjustment to random domain counts, and the modified regression estimators of Singh and Wu (1996, 2003) which use different working covariance for regression to obtain a calibration form of MF estimators. For the special case of simple random samples, the Fuller-Burmeister maximum likelihood estimator (Fuller and Burmeister, 1972), which generalizes both Lund (1968) and Hartley (1962, 1974) estimators, can also be expressed approximately in a GMHT-reg form. Details of the above generalization of the GMHT class to include regression estimators for MF surveys are considered in Singh and Mecatti (2014). They also considered a generalization of Kalton-Anderson estimator via modified-regression to obtain a calibration form within the COMB class. Finally, we note that a new interesting direction of development in MF estimation corresponds to the use of pseudo-empirical likelihood by Wu and Rao (2006) and Rao and Wu (2010) which takes a totally different approach and more research is needed to see how they can address the types of regression problems GMHT-reg can handle.

References

Bankier, M. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association*, 81:1074–1079.

Birnbaum, S. and Sirken, M. (1965). Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates. *Vital and Health Statistics*, Series 2, No. 11:65–82,149.

Ernst, L. (1989). Panel surveys. *John Wiley and Sons New York*.

Eurostat (2000). Push and pull factors of international migration, country report: Italy. Technical Report 5, Eurostat, Brukelles.

Fuller, W. and Burmeister, L. (1972). Estimators of samples selected from two overlapping frames. *Proceedings of the Social Statistics Sections, American Statistical Association*, pages 245–249.

Gonzales, G., Villalobos, A., and Wallace, M. (1996). Multiple frame agricultural surveys. *FAO Rome*, 1-2.

Hartley, H. (1962). Multiple frame surveys. proceedings of the social statistics sections. *American Statistical Association*, pages 203–206.

Hartley, H. (1974). Multiple frame methodology and selected applications. *Sankhya, Series C*, 36:99–118.

Iacan, D., Iachan, R., and Dennis, M. (1993). A multiple frame approach to sampling the homeless and transient population. *Journal of Official Statistics*, 9:747–764.

Kalton, G. and Anderson, D. (1986). Sampling rare populations. *Journal of Royal Statistical Society*, A:65–82,149.
Kwok, P., Shin, H., Murphy, W., O'Muircheartaigh, C., Debello, A., and Carris, K. (2009). Cross-community comparison and multi-frame weighting in reach u.s. *Proceedings of the Joint Statistical Meeting - Section on Survey Research Methods*, pages 2655–2662.
Lavallée, P. (2007). Indirect sampling. *Springer Series in Statistics*.
Lohr, S. (2009). Multiple frame survey in sample survey: Design, methods and applications. *Handbook of Statistics, Pfeffermann D. and Rao C.R. Eds*, 29 A:71–88.
Lohr, S. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Survey Methodology*, 37:197–213.
Lohr, S. and Rao, J. N. K. (2006). Estimation in multiple frame surveys. *Journal of the American Statistical Association*, 101:1019–1030.
Lund, R. (1968). Estimators in multiple frame surveys. *Proceedings of the Social Statistics Sections, American Statistical Association*, pages 282–288.
Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, 33:151–158.
Rao, J. (1968). Some nonresponse sampling theory when the frame contains an unknown amount of duplication. *Journal of the American Statistical Association*, 63:87–90.
Rao, J. N. K. and Wu, C. (2010). Pseudo empirical likelihood inference for multiple frame surveys. *Journal of the American Statistical Association*, 105:1494–1503.
Singh, A. and Mecatti, F. (2009). A generalized multiplicity-adjusted Horvitz Thompson class of multiple frame estimators. *Book of Abstract - ITACOSM09, June 10-12/2009, Siena, Italy*, Invited Paper:75–77.
Singh, A. and Mecatti, F. (2011). Generalized Multiplicity-adjusted Horvitz-Thompson type Estimation as a Unified Approach to Multiple Frame Surveys. *Journal of Official Statistics*, 27:633–650.
Singh, A. and Mecatti, F. (2014). Use of zero functions for combining information from multiple frames. *in Contributions to Sampling Statistics - ITACOSM 2013 Selected Papers*, F. Mecatti, P.L. Conti and M.G. Ranalli Editors, Springer, 2014. See also Singh, A.C. and F. Mecatti. Generalized Multiplicity-Adjusted Multiple Frame Estimation via Regression. Proceedings of Statistics Canada Symposium 2013 Producing reliable estimates from imperfect frames.
Singh, A. and Wu, S. (1996). Estimation for multiframe complex surveys by modified regression. *Proceedings of the Statistical Society of Canada, Survey Methods Section,n*, pages 69–77.
Singh, A. and Wu, S. (2003). An extension of generalized regression estimator to dual frame surveys. *Proceedings of the Joint Statistical Meeting - Section on Survey Research Methods*, pages 3911–3918.
Sirken, M. G. (2004). Network sample surveys of rare and elusive population: a historical review. *Proceedings of Statistics Canada Symposium*.
Skinner, C. and Rao, J. (1996). Estimation in dual frame surveys with complex designs. *Journal of American Statistical Association*, 91:349–356.
Skinner, C. J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *Journal of the American Statistical Association*, 86:779–784.
Wu, C. and Rao, J. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *Canadian Journal of Statistics*, 34(3):359–375.