



Global Strategy
IMPROVING AG-STATISTICS



Improving methods for linking area frames with list frames: preliminary results



Cristiano Ferraz

Federal University of Pernambuco, UFPE



BRIEF DESCRIPTION OF THE RESEARCH TOPIC

- Review the literature about LIST FRAMES, AREA FRAMES, and DUAL FRAME ESTIMATION
- REQUIREMENTS, ADVANTAGES and DISADVANTAGES
- Search for an ideal scenario for producing AGRICULTURAL STATISTICS





LITERATURE REVIEW (1)

LIST FRAMES (1)

- [5] Carfagna E. and Carfagna A. (2010). *Alternative sampling frames and administrative data; which is the best data source for agricultural statistics?* . in Benedetti, Bee, Espa, Piersimoni (Editors), *Agricultural Survey Methods*, Wiley, Chichester, UK.
- [15] Lavallee, P. (2005). Quality indicators when combining survey data and administrative data. In *Proceedings of Statistics Canada Symposium 2005: Methodological Challenges for Future Information Needs*. <http://www.statcan.gc.ca/pub/11-522-x/11-522-x2005001-eng.htm>..



LITERATURE REVIEW (2)

LIST FRAMES (2)

Advantages and disadvantages

- Usually a list of holdings or holders addresses
- Often incorporate auxiliary information
- Degenarates quickly over time
- Needs maintenance



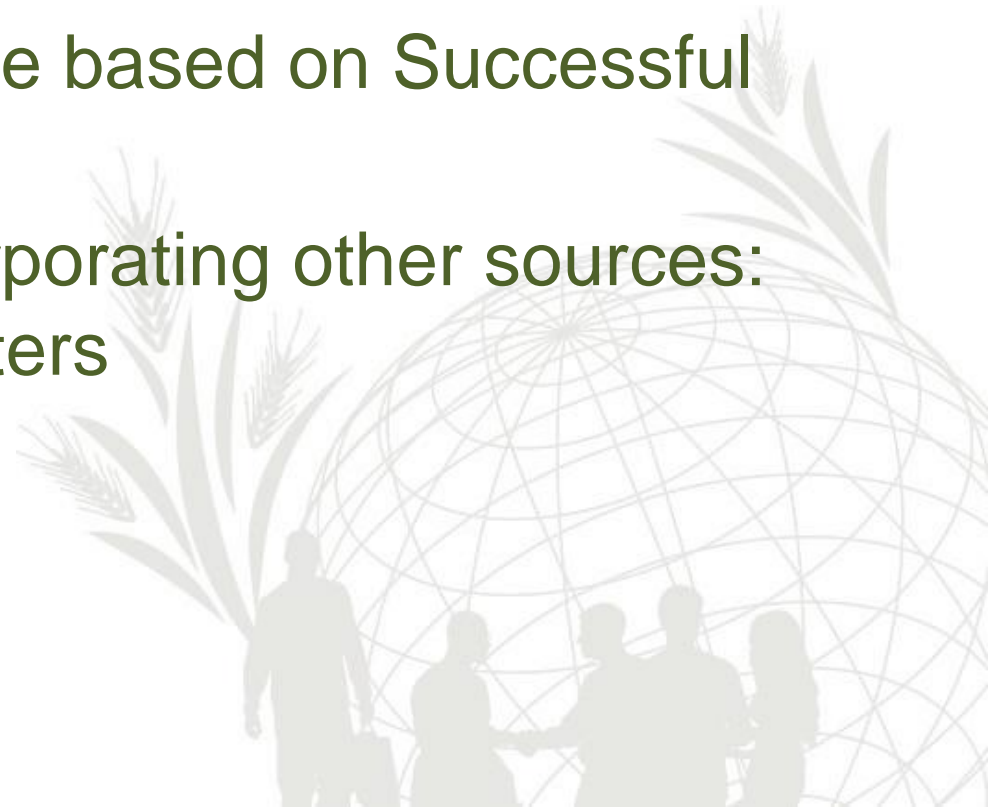


LITERATURE REVIEW (3)

LIST FRAMES (3)

Complete Coverage, No Duplication, Updated

- Building a List Frame based on Successful Censuses
- Advantages of incorporating other sources: administrative registers





LITERATURE REVIEW (4)

AREA FRAMES

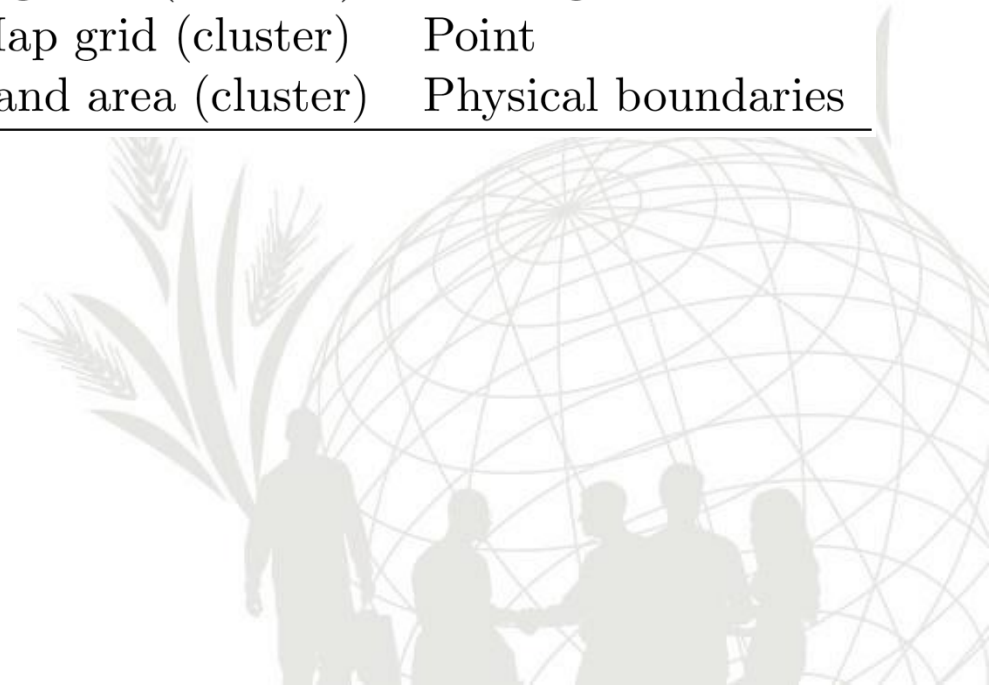
Complete Coverage, No Duplication, Remains Updated for a long time

- [7] FAO. (1989). *Sampling Methods for Agricultural Surveys*. FAO Statistical Development Series No. 3. Rome, FAO.
- [8] FAO. (1996). *Multiple Frame Agricultural Surveys, Volume 1*. FAO Statistical Development Series No. 7. Rome, FAO.

LITERATURE REVIEW (5)

Table 1: types of area and list frame suitable for agricultural surveys

Frame type	Frame description	Unit component	Unit type example
1	List frame	Element	Holder addresses
2	List frame	Cluster	Villages
3	Area frame	Segment (element)	Holding area
4	Area frame	Map grid (cluster)	Point
5	Area frame	Land area (cluster)	Physical boundaries





LITERATURE REVIEW (6)

DUAL FRAME DESIGN

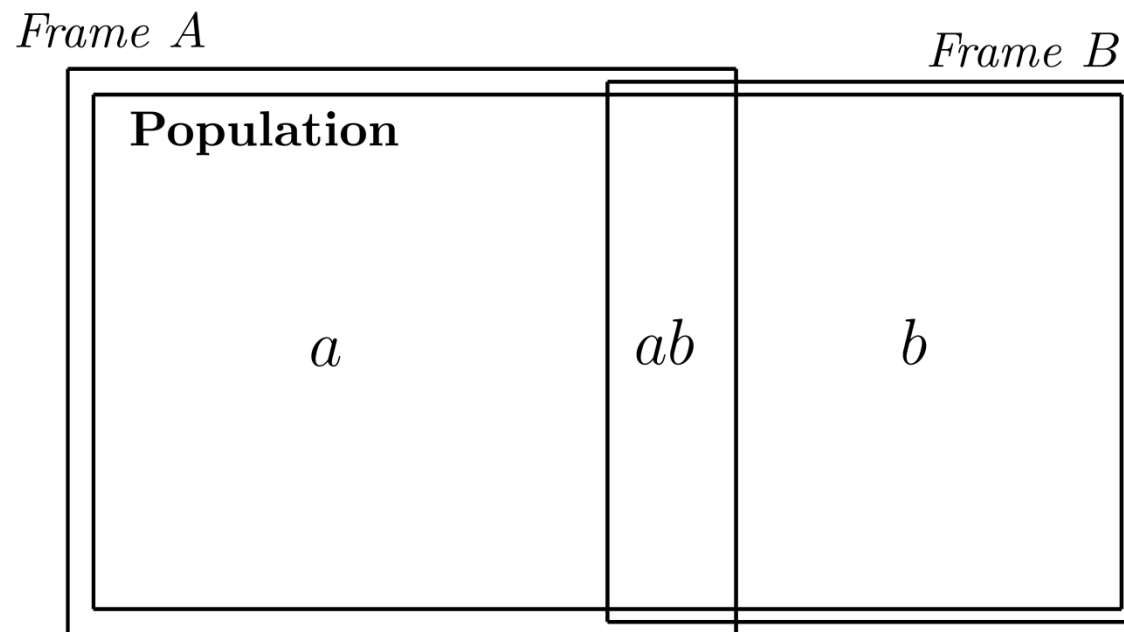


Figure 1: General dual frame scenario

LITERATURE REVIEW (7)

DUAL FRAME DESIGN: Area frame and List frame

Frame A

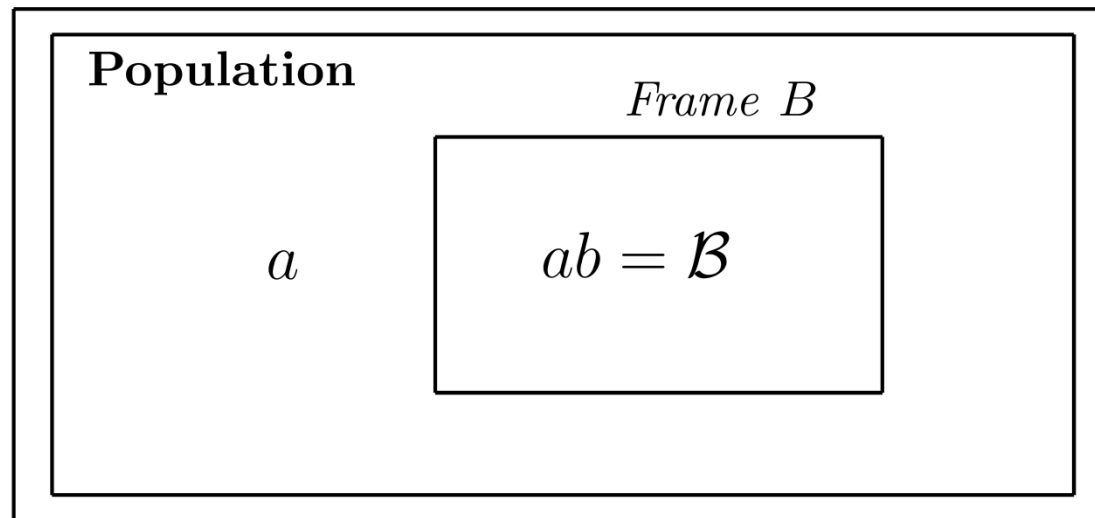


Figure 2: Special case dual frame scenario based on area frame and list frame



LITERATURE REVIEW (8)

Dual Frame Design Estimation

$$\hat{t} = \hat{t}_a + \hat{t}_b + \hat{t}_{ab}$$

$$Var(\hat{t}) = Var(\hat{t}_a) + Var(\hat{t}_b) + Var(\hat{t}_{ab}) + Cov(\hat{t}_a, \hat{t}_{ab}) + Cov(\hat{t}_b, \hat{t}_{ab})$$



LITERATURE REVIEW (9)

Hartley

$$\hat{t}_H = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} y_k^*$$

$$y_k^* = \begin{cases} p \frac{y_k}{\pi_k^A}, & \text{if } k \in S_{ab}^A \\ (1 - p) \frac{y_k}{\pi_k^B}, & \text{if } k \in S_{ab}^B \end{cases}$$



LITERATURE REVIEW (10)

Fuller and Burmeister

$$\hat{t}_{FB} = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} y_k^* + p_2 \left(\hat{N}_{ab}^A - \hat{N}_{ab}^B \right)$$

$$y_k^* = \begin{cases} p_1 \frac{y_k}{\pi_k^A}, & \text{if } k \in S_{ab}^A \\ (1 - p_1) \frac{y_k}{\pi_k^B}, & \text{if } k \in S_{ab}^B \end{cases}$$



LITERATURE REVIEW (11)

Bankier

$$\hat{t}_B = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} \frac{y_k}{\pi_k^A + \pi_k^B}$$





LITERATURE REVIEW (12)

Skinner and Rao

$$\hat{t}_{SR} = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} \gamma y_k^*$$

$$y_k^* = \begin{cases} \frac{y_k}{\pi_k^A}, & \text{if } k \in S_{ab}^A \\ \frac{y_k}{\pi_k^B}, & \text{if } k \in S_{ab}^B \end{cases}$$

$$\gamma = \frac{\hat{N}_{ab,SR}}{\left(\pi_k^A \hat{N}_{ab}^A + \pi_k^B \hat{N}_{ab}^B \right)}$$



LITERATURE REVIEW (13)

Record Linkage (1)

$i \in A$ $j \in B$ pair (i, j)

$(i = j)$ $i \neq j$

D can assume values d_1, d_2 or d_3

LITERATURE REVIEW (14)

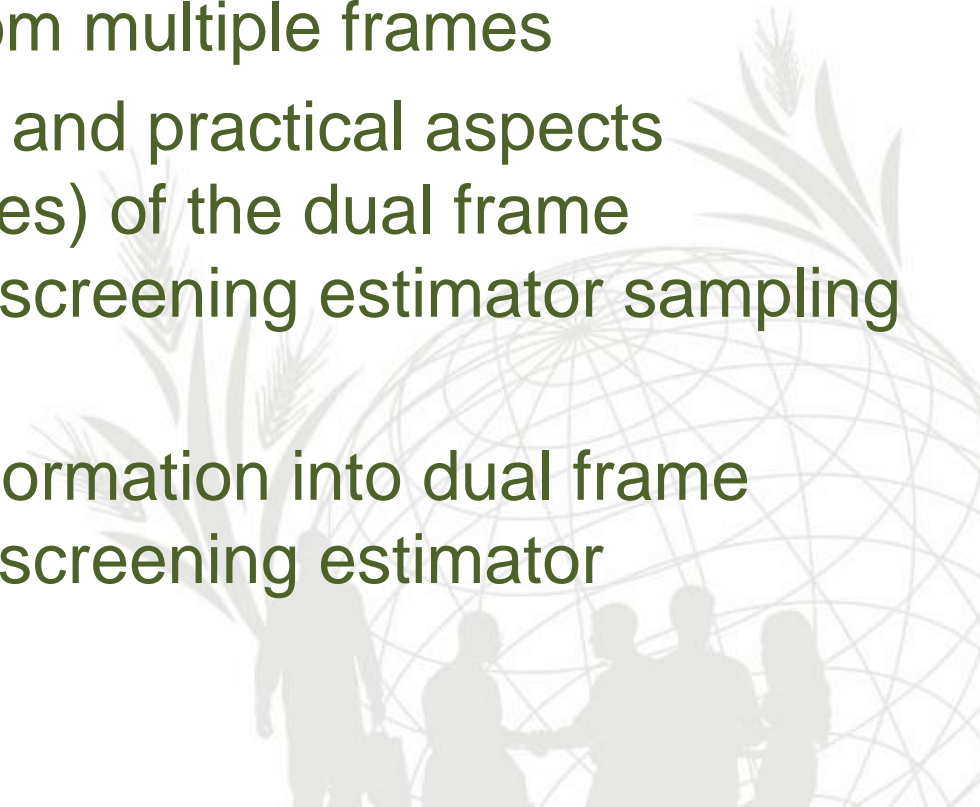
Record Linkage (2)

- Type I Error: $D(i, j) = d1 \mid d3$ is true;
- Type II Error: $D(i, j) = d3 \mid d1$ is true.

$$R = \frac{P(D(i, j) \mid (i, j) \in M)}{P(D(i, j) \mid (i, j) \in M^c)} = \frac{p_{ij}}{p_i p_j}$$



SUB-TOPICS REQUIRING FURTHER RESEARCH

- Optimum size of list frames in a dual frame approach
 - The use of permanent random numbers for coordinating sampling from multiple frames
 - Comparing the efficiency and practical aspects (related to available frames) of the dual frame estimators, including the screening estimator sampling from multiple frames
 - Incorporating auxiliary information into dual frame estimators, including the screening estimator
- 



THANK YOU

(Cristiano Ferraz, UFPE, cferraz@de.ufpe.br)

