geographical level. The problematic: Are Small Area Estimation achieving more accuracy for estimating population or not? And what about evaluating of SAE for geographical hierarchy?

- To answer this problematic the paper will apply two methods of small area estimation: Direct Estimation Methods such as direct estimator and broad area estimator without auxiliary data, and Indirect Estimation Methods such as unit area level by using binary logistic regression. - Small area estimation methods will be applied on LFS 2006, Census 2006 to estimate unemployment rate, therefore the choice of the used method depends on the available data as well as the specification of the required domain estimation, so this paper will be applied on Qaliobia governorate that reflects the nature of Egyptian society. This paper uses two methods of small area estimation to give unemployment estimates at different small areas across the geographical hierarchy (i.e.Shiakha\village) to know the accuracy of SAE methods and comparing them with LFS estimates for choosing which methods will be more accurate.

2- **Relevant literature and researches (previous studies): -**

The Canadian experience used a number of primary sources of data: the Canadian census of population and housing, the Canadian LFS and the Federal Governorate Unemployment Insurance (UI). Other small area labor market data were used as auxiliary information. The main goal was to minimize the model estimation error, so they tested three estimation techniques: synthetic estimation, SPREE (Structure Preserving Regression Estimation) and regression estimation. The Canadian experiment showed that there are highly correlated variables between data sources, despite basic differences in definition. -United Kingdom: Office for National Statistics (ONS) The ONS in the UK, in conjunction with Southampton University, conducted research about the estimation of unemployment based on the ILO definition. A number of different approaches were explored utilizing the high correlation between the LFS estimates of unemployment and the number of job seekers allowance benefit claimants. The main approach was to develop regression models linking the unemployment estimates to the claimant count information by local authority district (ONS, 2001b). - Hakizimana,J.(2011) showed the application of the small area estimation methods for employment status ,and resolving the problem of the misclassification of employment status in Census 2001 by readjusting the data; In addition to respect the classification of people as employed, unemployed or economically inactive to that of Labor Force Survey of September 2001. This research investigates the use of these SAE methods to provide better estimates of employment status at a small area (municipal) level, and reducing the gap classification between who have jobs and those who do not in census for LFS.

3- **Methodology:-**

Conducting this research depends on:
 Data Sources:
 This research used two sources of data: 2006 Census Data (population) and 2006 LFS data (sample).
The difference between SAE Methods depends largely on the available data. The choice of SAE methods depends on the availability of auxiliary data, the relationship between these data and the variables of interest at the small area level, to get the best model for estimation and the methods for SAE include direct and indirect estimators. Direct estimators use only data from the small area being examined. Indirect estimators are better due to use observations from related variables of domain .the methods divided to:
The choice of SAE methods depends on the availability of auxiliary data, the relationship between these data and the variables of interest at the small area level This research used two sources of data: 2006 Census Data and 2006 LFS data .The applied SAE Methods are:
1-   The Direct Estimation methods:
   1-1   Direct Estimator: Appling the weights of LFS at Shikhia \ Village level.
   1-2   Broad Area Ratio Estimator with no auxiliary data(BARE):

It is calculated by applying the rate for a broad area obtained from a survey (e.g. unemployment rate) to the small area populations. The success of the BARE hinges largely on the choice of the broad area. And this method depends on this assumption: - The population characteristics of interest are homogeneous for all small areas within a broad area, and sample allocated represents the characteristics of that small area. Appling direct estimation Broad Area Ratio for each small area inside broad area (Kism/Markaz).

2- Indirect Estimation Methods :

2-1 Unit level estimator (Logistic Regression Model - LRM).

This research will use two methods (Direct and Indirect methods) with three estimations (direct, broad area ratio without auxiliary data and unit level (binary logistic regression)) to estimate unemployment rate in Egypt for small areas (Shiakhias, Villages), using Census and LFS for direct and indirect methods of SAE. Evaluating these estimates to show accurate estimates, because the problem of SAE is how to produce reliable estimates of characteristics of interest such as means and rates.

Independent variables	Categories of independent variables (dummy variables)
Eduction status	edu_1 refer to Clueless or less of age, edu_2 refer to Reads and writes, edu_3 refer to Lower than the average qualified , edu_4 refer to average qualified, edu_5 refer to Qualified above average, edu_6 refer to bachelor's degree, edu_7refer to qualified above bachelor's degree
Gender	Male and female
Place of residence	Urban and rulur
Age	Age_1 refer to Less than 15 years, Age_2 refer to 15-29 years, Age_3 refer to 30-44years, Age_4 refer to 45-64 years
Marital Status	Married and not Married
Dependent variable Working Status	Employed and unemployed (dummy variables with 1 to unemployed and 0 to employed)

4- Results

Table1: The Evaluation of SAE for two SAE Methods:

Governorate	Urban / Rular	Kism/ Markaz	Shikhia/ Village	Unemployment rate of Census with direct estimation	Unemployment rate of LFS with direct estimation	Unemployment rate of Census with BAE	Unemployment rate of LFS with BAE	Unemployment rate of Census with LRM	Unemployment rate of LFS with LRM
Qaliobia	urban	qism banha	kafar manaqir	5.57	2.34	5.9	8.47	14.86	8.11
	rulur	markaz alkhanka	23 yuliu	3.21	10.96	5.9	8.47	14.45	12.68

The pervious table shows the evaluation of SAE with example for 2 shiahia\ Village from 40 in Qaliobia.

Figure (1) Comparison between the estimation methods of unemployment rate for LFS 2006

- The independent variables (gender, age, education status, Place of residence and marital status) explained 70.5% of the variation in dependent variable (working status).
- After including model of independent variables the classification of the model increased from 84.9% to 94.3% to classify of working status and reflect that the independent variables in model have strong effect in working status and they are significant variables.
- <u>Direct Estimator</u> showed the difference between the measures by the LFS and the Census which is due to the misclassification working status in Census, So the unemployment rate in LFS is more reliable than Census data.
- <u>BARE</u> showed that the unemployment rate in each small area (shiakha's/ villiages) is 8.47% and it is nearly of the direct estimation of the governorate level (Qaliobia) as a broad area is 8.88% that reflects reliable accuracy of LFS. For Census's unemployment rate of Qaliobia is 9.4% and BARE for Census is 5.9%, that reflects not reliable enough because the same problem of misclassification in Census.
 - Unit level estimation with LRM (Logistic Regression Model) as Indirect Estimation depends on the allocation between Census and LFS properties to solve misclassification of employment status in Census and small sample size in LFS for shaikhia's / village's with this model. There is little remarkable difference between LFS estimates and the Unit Level Estimator, and larger changes of unemployment rates of Census have been increased by applying this model. The LRM estimation from LFS has increased of governorate level from 8.8% for LFS to 9.8% by applying this model (LRM), and the Census data had the unemployment rate 10.9 (Census' results) and increased to 13.2% after applying LRM model on Census's data, because that have problem's classification in unemployment status that reflect the efficiency of model of estimation; and also the efficiency of small area estimation.
- The result of using the Indirect Estimation (Logistic regression Model) has reduced the gap between the Census and the LFS profiles of unemployment status. This is an indication that the estimation has borrowed strength in the estimation of unemployment status from the LFS and applied this to the Census.
- The RRMSE is small (Relative Root Mean Square Error is primarily a measure of prediction error) of unemployment estimated rates of LRM for all (shiakhia/ village) in Qualiobia governorate. It is less than 25% with large percentage that reflect the accuracy of these estimates with logistic Regression Model.
- Aggregation of small area estimates to larger regions such as shiakhias/ villages to governorate level and compare them with LFS results. It will improve the accuracy of the estimates and explain that LRM as Indirect Method is more accurate than other estimators such as Direct and BARE of Direct Method, because this method depends on the statistical model for Census and LFS ; so it is recommend to apply this method in other researches Such as Poverty rate and Disability rate in Egypt .

5- Conclusions

- The small area estimates are tools. Used in conjunction with understanding of local area characteristics and their reliability limitations, they should assist in making sensible decisions on issues involving the regional distribution of unemployment rates .This research showed the concepts and definitions of small areas as well as concepts of unemployment. The importance of small area data is to support planning, decision making and service delivery at a local area level requires statistics on the smallest geographic level. The goal of SAE is the availability of unemployment rates on small geographic level in the light of poor quality classification of working status in Census questions and good classification is done through Labor Force Survey (LFS) questions. LFS has a problem due to small sample size in this geographical level then estimates inaccurate, therefore it has been replaced by using the characteristics of the census for such a small geographical levels to treat the small sample size to get reliable estimates of small areas. The SAE methods include direct estimation method (direct estimator and Broad Area Ratio Estimator (BARE) without auxiliary data) that depends on survey estimates, BARE assuming that estimates from the small area (i.e. Shiakha or Village) have the same characteristics as the large area (i.e. governorate) ,and these estimates might be biased due to the assumption of homogeneity between large and small areas not being valid. Indirect Estimation method that can be used (unit area level estimation by Binary Logistic Regression Model (LRM), which depends on the statistical models using the data Census as auxiliary information for demographic characteristics to derive estimates. - So this research focuses on solving the problem of measuring unemployment rate at the smallest geographical level (Shiakha / Village) in Egypt with more accuracy and details to get the main reasons of this problem. Which help policymakers to address this national phenomenon at this small geographical level?
- The results for this research is estimating unemployment rate on the smallest geographical level (shiakha / village) in the Republic using indirect method was more accurate than direct method due to using statistical model (binary logistic regression) with auxiliary data from Census and LFS .
- So this research helping policymakers to address unemployment problem at small geographical level to make sure of the efficiency of this estimates and depending on them to make decisions, draw policies and identify the gap between employed and unemployed across the geographical hierarchy (i.e. Shiakha\Village), and helping to solve the unemployment problem without need to big sample size to study the problem and solve it. This research recommends with applying SAE in other phenomenon such as Poverty, Disability rate ….etc .

- This paper focuses on solving the problem of measuring unemployment rate at the smallest geographical level (Shiakha / Village) in Egypt with more accuracy and details to get the main reasons of this problem. Which help policymakers to address this national phenomenon at this small geographical level?
- The results for this research is estimating unemployment rate on the smallest geographical level (shiakha / village) in the Republic using indirect method was more accurate than direct method due to using statistical model (binary logistic regression) with auxiliary data from Census and LFS .
- So this research helping policymakers to address unemployment problem at small geographical level to make sure of the efficiency of this estimates and depending on them to make decisions, draw policies across the geographical hierarchy (i.e. Shiakha\Village), and helping to solve the unemployment problem without need to big sample size to study the problem and solve it.

References:

1- Australian Bureau of Statistics (2005), Information Paper: Regional Research in Australia the Statistical Dimension: an Information Development Plan for Rural and Regional Statistics, Catalogue No. 1362.0 (Due for release in early 2006).
2- Brackstone,G.J.(1987) .Small area data : policy issues and technical challenges.In small area Statistics(R.Platek,J.N.K.Rao,C.E.Sarndal and M.p.Singh eds.)3-20.wiley,New York.
3- Brown, G., Chambers, R., Heady, P. and Heasman, D. (2001), Evaluation of Small Area Estimation Methods – An Application to Unemployment Estimates from the UK LFS, Proceedings of Statistics

Canada Symposium 2001,Achieving Data Quality in a Statistical Agency: A Methodological Perspective.

4- Francisco J. (2003), "Small area estimation: An overview of existing methodologies with application to the Estimation of unemployment rates in the Philippines", University Laval, Canada.

5- Hakizimana,J.(2011)," Small Area Estimation Of Unemployment For South African Labor Market Statistics ", Faculty of science.

6- Khoshgooyanfard, A.R. and Monazzah, M.T., (2006), "A Cost Effective Strategy for Provincial Unemployment Estimation: A Small Area Approach", Survey Methodology, Vol. 32, No. 1, 105-114, Statistics Canada.

7- Office for National Statistics ONS (2001a), "Small area estimation project", London, UK.

8- Rao J.N.K, (2003a), "Small Area Estimation", John Wiley & Sons, Inc., New York.

9 - Saei, A and Chambers, R. (2003) Small Area Estimation: Methods Based on the Application of Mixed Models, Southampton Statistical Sciences Research Institute Methodology Working Paper M03/16

10- Satorra A. and Ventura E, (2006), "Small Area Estimation at IDESCAT: Current and Related Research", IDESCAT, Barcelona, Spain, 22 pages.

11- Australian Bureau of Statistics (2005), A Guide to Small Area Estimation, available online at http://www.nss.gov.au/nss/home.NSF/pages/Small+Areas+EstimatesOpenDocument,   last accessed on 17/09/07.

12-Labor Force Survey bulletin 2006 by CAPMAS at web site http://www.capmas.org.eg/