

Calibration in Transition to New Urban-Rural Definitions in LFS

Hasan Cenk Dedeođlu*, Cansu Öztürk, Dr. Duygu Kılıç

cenk.dedeoglu@tuik.gov.tr

Introduction

In this study, it was aimed to examine the effects of this situation on the indicators of the survey if new urban-rural definitions are used in the calibration of the household labour force survey. In the first part of the study, brief information about the history and survey design of labour force survey in Türkiye is given. In the second part, past and present urban-rural definitions used in Türkiye are introduced briefly. In the third part, the studies carried out to measure the effects of using the new urban-rural definition in the calibration of the labor force survey on the estimates are mentioned and the findings are presented. Finally, the study and comments are summarized in the conclusion section.

1. History and Survey Design of Household Labour Force Survey in Türkiye

Turkish Statistical Institute (TurkStat) conducted the first Household Labour Force Survey (LFS) in 1966. Then, it was applied at irregular intervals until 1985. Since 1988 LFS has been implemented regularly by TurkStat. In 1995, the LFS field application was started to be carried out with the Computer Assisted Personal Interviewing (CAPI) method, and the survey results had been published by Türkiye, urban and rural areas.

In the year 2000, some important changes were made on the survey application period, sample size, estimation dimension and questionnaire of the survey. In this context, monthly surveys have been implemented since 2000, and the survey results were published quarterly between 2000-2003 on Türkiye, urban and rural basis and annually on the basis of 7 geographical regions and 9 selected provincial centers. Since 2004, the results of the survey have been published annually according to the Classification of Statistical Regional Units (NUTS) Level II. From 2005 to 2020, the estimates of the labour force survey were published in every month based on the three months moving average results.

In 2014 continuous survey application was started in which not only the first week of the month but all weeks of the year were defined as the reference period. While the period to be taken as a basis for job search, which is one of the criteria in the measurement of unemployment, was determined as "3 months" before 2014. This period started to be accepted as "4 weeks" in accordance with the relevant European Union (EU) regulations. In addition, the indicators started to be calculated according to the revised non-institutional population estimates based on the 2014 administrative division determined according to the Metropolitan Law of 6 December, 2012. As of January 2020, in addition to the CAPI data collection method, Computer Assisted Telephone Interviewing (CATI) method has been applied and mixed-mode data collection method is started to be implemented. In order to better measure the changes that have occurred in the structure of the labour market over the years, new "definitions" and "standards" for the measurement of employment and unemployment were determined with the 19th ICLS carried out by the ILO in 2013. Starting from 2021, new definitions and concepts have been started to be implemented in the LFS as per the relevant EU regulation. In addition, improvements were made in the publishing method in order to share the produced estimates with users more timely. Accordingly, independent monthly estimates for main indicators and quarterly estimates for detailed labour force statistics have been started to be published instead of three months moving average estimates. After the editing, the comparable time series of the main labour force indicators were calculated retrospectively until 2005 using statistical and econometric methods.

The LFS is a survey conducted with the sampling method. The sampling unit is the household address and the observation unit is the household. All private households who are residing in the territory of Republic of Türkiye are covered. Residents of schools, dormitories, kindergartens, rest homes for elderly persons, special hospitals, military barracks and recreation quarters for officers are excluded from the sampling frame. Within the scope of the LFS, the survey is conducted with the household resides at the selected sample address. The sample size of the survey has been determined as 58,590 households quarterly and 234,360 households (approximately 0.2% sampling rate) annually in order to meet the requirements of The Regulation No. 2019/2240 of EU. Households to be interviewed are selected randomly providing a representation of Türkiye per month and quarter, and of NUTS Level II per year by using two-stage stratified cluster sampling method. The LFS was designed as a semipanel survey based on address follow-up in order to better measure the changes between the consecutive quarters and the consecutive years. A rotation pattern formed to ensure 50% overlap between two consecutive quarters and the same quarters of two consecutive years for selected samples. According to this sampling design, a household that participated in the sample is interviewed four times (waves) in a 15-month period. There are 13 weeks in each quarter of the study and the sample size has been equally distributed to the weeks in each term.

Weighting is a method used to obtain representative estimates of population parameters from the sample. While reaching the final weight for LFS, the design weights are calculated according to selection criteria and then non-response adjustment, and external distribution controls and trimming are applied. Information on age group x gender, NUTS 3 Level (81 provinces), NUTS 2 Level (26 regions) x urban-rural and household size is used as an external distribution [1].

Design weights are calculated by the inverse of the selection probabilities. In the LFS, the design weight is $w_{h_i} = \frac{1}{p_{h_i}}$.

Non-response adjustment is applied at individual, household and block (cluster) levels. Then final weights are calculated by calibrating adjusted weights to projected population distributions using Integrative Calibration Method. It is necessary to use external auxiliary variables (if available) to increase the precision of selected sample while producing estimations at national level. The use of external available datasets is especially necessary when the sample size is not enough for small areas, high nonresponse rate is occurred or the results of the sample lacks for producing correct estimations on certain characteristics of the population. Therefore, the external datasets used are mainly the marginal distribution of some characteristics of the population. Geographical urban - rural, age group or gender distributions can be shown as example.

4 different marginal distributions are used in the calibration process for LFS since 2021;

- Türkiye Age group - Gender distribution,
- NUTS3 total distribution,
- NUTS2 x Urban - Rural distribution,
- Türkiye household size distribution.

2. History of Urban-rural Studies at TurkStat

Defining areas that are economically and functionally rural or urban based solely on administrative and/or legal status is insufficient in statistical production and policy formulation. There was a need to revise the definitions made in this way in order to provide more meaningful input to policy makers and to reveal the urban-rural structure that changes with urbanization. For this reason, the definitions of rural and urban that have been used in the TurkStat since the beginning of 2014 have been revised.

According to the urban-rural definition based on population and legal boundaries used by TurkStat at the beginning of 2014, 28% of Türkiye was seen as rural and 72% as urban. After the urban-rural definitions were changed in accordance with the laws, 14% of Türkiye's population became rural and 86% urban.

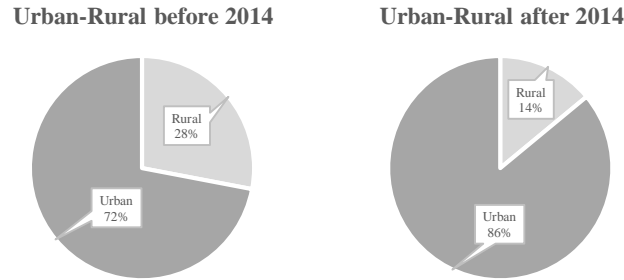


Figure 1: Change of Urban-Rural Rate

The fact that this definition depends only on legal borders and population makes it difficult to produce rural development indicators and statistics at the urban-rural level, negatively affecting the decision-making mechanisms of policy makers. Taking these negativities into consideration, it is aimed to develop urban-rural definitions that are in line with European Union (EU) standards and independent of administrative borders, taking into account both national and international needs.

In the studies of producing a urban-rural definition, the methods in the booklet "Applying the Degree of Urbanization - A methodological manual to define cities, towns and rural areas for international comparisons" published by Eurostat in 2021 were applied (Eurostat, 2021). In the Degree of Urbanization (DEGURBA) methodology, a urban-rural class assignment is first made at the grid level. After the grid classes are determined, settlement classification is started.

While establishing the urban-rural definition specific to Türkiye, the DEGURBA approach determined by Eurostat, which defines the degree of urbanization of an area at two different levels, triple and seventh, was taken as basis. The Degree of Urbanization is based on 1 km² population density grids. Address-Based Population Registration System results and Spatial Address Registration System are the basic data sources, and as a result of their matching, coordinated population (household population) data is obtained at the grid level. This information was moved from the grid level to the administrative borders to ensure continuity in TurkStat's current statistical production process. Since the provincial and district levels are insufficient to reflect the urban-rural characteristics, it was deemed appropriate to make urban-rural definitions in three classes at the neighborhood-village level.

According to the newly created urban-rural definition (DEGURBA Level 1) 17,7% of Türkiye's population is in thinly populated areas, 67,3% is in densely populated areas, and 15% is in intermediate-density areas [2].

3. Using DEGURBA in LFS

The calibration variables currently used in the Household Labor Force Survey are given below;

- 1) Türkiye Age group x Gender distribution,
- 2) NUTS3 total distribution,
- 3) NUTS2 x DEGURBA Level 1 distribution,
- 4) Türkiye household size distribution.

In this study, the effect of using DEGURBA definitions on the estimations was examined. All other calibration variables were kept constant and only NUTS2 x DEGURBA Level 1 was used instead of NUTS2 x urban-rural.

To see these effects, twelve monthly data was considered. Using this data, the unemployment rate, employment rate and labor force rate variables, which are among the main labor force indicators, were calculated both in total and by gender. In addition, time-dependent underemployment and potential labor force variables were also examined. In the total level examination without breakdowns, it was observed that the maximum difference is around 3% for Türkiye. As an example of this situation, the change in the unemployment rate is given in Figure 2. Months are randomly ranked from 1 to 12 according to the magnitude of change.

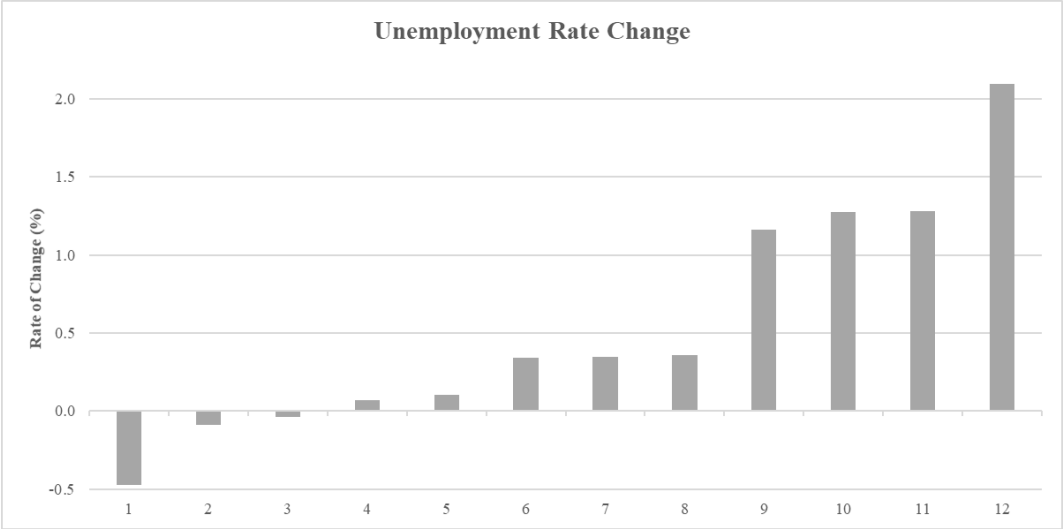


Figure 2: Unemployment Rate Change

When the annual data was analyzed by sector, it was seen that the most striking sector was agriculture. In Figure 3, the variation in estimates for the agricultural sector is given in the NUTS2 regional breakdown. Here, the regions are randomly ranked from 1 to 26 according to the magnitude of change. They are not listed in the order of publication. When Figure 3 is examined, it is seen that there are serious differences in the agricultural sector estimates obtained after the transition to DEGURBA.

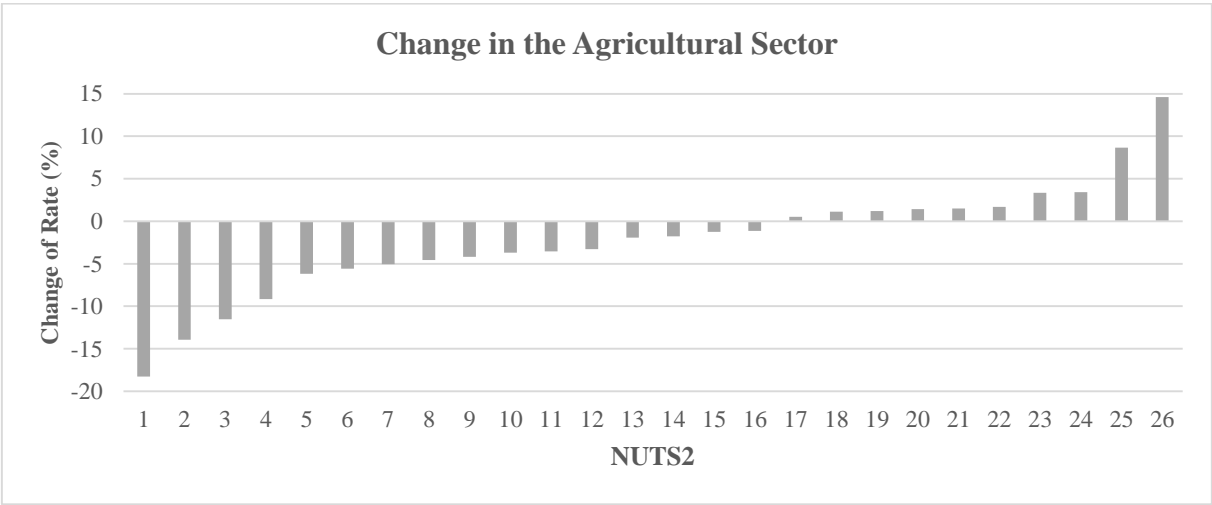


Figure 3: Change in the Agricultural Sector

4. Conclusion

In this study, the variation in the LFS estimates created by the change in the urban-rural variable, which is one of the calibration dimension, was examined both in the country total and in breakdowns such as regional, sectoral and gender discrimination.

After the urban-rural definition change in 2014, it was observed that there was a significant decrease in rural areas especially in 30 provinces called metropolitan cities. For these reasons, although urban-rural distinction was used only in the sampling design, it could not be used as an estimation domain. Therefore, TurkStat stopped publishing estimates on a urban-rural basis in 2014. As a solution to this problem, DEGURBA definitions were introduced after the Spatial Address Registration System established in our country. After the studies conducted on DEGURBA methodology, it was seen that the DEGURBA provides more meaningful information in terms of defining urban and rural areas. That is why it was decided to use urban-rural breakdown both at survey design and at producing estimates on this basis. Since the use of DEGURBA in household sample surveys came to the fore after this decision, it became necessary to revise the sampling design of surveys and measure the effect of the use of new definitions on the statistics produced. For this purpose, sample sizes and calibration dimensions revised according to DEGURBA definitions. To be able to see the possible effect on estimates beforehand, analyses done on previous years' data and estimates were recalculated. Differences between point estimates and CV values of the estimates were examined to show the effect on estimates and the quality of the statistics.

In conclusion, the projected population distributions based on the new definition were obtained, then the respondent LFS data was re-coded according to new urban-rural definition. Afterwards, the data entered the calibration process with these reclassification and new projected populations. The previous twelve monthly independent main estimates (Labour force, Employment, Unemployment, Not in labour force) were compared in total level and seen that the maximum difference is around 3%. Agricultural sector estimates were also examined based on yearly data since it was thought that the change in the definition of rural area would affect this sector.

It should be noticed that this is an indirect approach and the analysis were made on main indicators not on the more detailed indicators. In light of these developments, it is planned to restart publishing estimates on a urban-rural (DEGURBA level 1) breakdown in the coming years.

References

- [1] Methodological explanation document. Available at: <https://data.tuik.gov.tr/Bulten/Index?p=Isgucu-Istatistikleri-2022-49390#:~:text=Zamana%20ba%C4%9Fl%C4%B1%20eksik%20istihdam%2C%20potansiyel,%17%2C%20olarak%20ger%C3%A7ekle%C5%9Fti.>
- [2] Urban-Rural Population Statistics Press release. Available at: <https://data.tuik.gov.tr/Bulten/Index?p=Urban-Rural-Population-Statistics-2022-49755&dil=2>