RESEARCH

JOURNAL OF STATISTICS AND SOCIO-ECONOMIC ANALYSES

1

INSTAT
Instituti i Statistikave

No. 01 / June 2018

This magazine is an independent periodical publication of the Statistics Institute.

INSTAT
Republika e Shqipërisë
Instituti i Statistikave

agenda
INSTITUTE

RESEARCH

JOURNAL OF STATISTICS AND SOCIO-ECONOMIC ANALYSES

INSTAT

This journal is published by INSTAT in collaboration with AGENDA Institute, with the support of the Swiss Agency for Development and Cooperation through their project PERFORM – Performing and Responsive Social Sciences, implemented by Helvetas and University of Fribourg

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

PERFORM
Performing and Responsive Social Sciences

HELVETAS

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

EDITORS

INSTITUTI I STATISTIKAVE
Blv. "Zhan d'Ark" , Nr. 3, Tiranë / Shqipëri
Tel: + 355 4 2222411 / 2233356
e-mail: Revistakerkimi@instat.gov.al
www.instat.gov.al

CONTENT

SCIENTIFIC ARTICLES

INFORMATIVE ARTICLES

Dear Readers,

Dear Readers,

INSTAT, as the main provider of official statistics, plays a crucial role on promoting sustainable statistical information empowerment on decision making, and encouraging the dialogue between producers of official statistics and researcher's community as well.

In order to help the audience to make better use of statistical information, INSTAT decided to put efforts in producing more detailed analyses on socio-economic phenomena and statistics through setting up an own scientific journal.

National Institute of Statistics takes the pleasure in presenting the first issue of journal "The Research", Journal of Statistics and Socio-Economic Analysis.

"The Research" is the first open access journal with the purpose to create a relevant and challenging publication that will serve as a gathering space for official statisticians, academics, graduate students, analysts, researchers, public policy makers, and in general, any party interested in the way in which statistical thinking influences the decisions that affect the different aspects of economy and society.

"The Research" will be published twice annually, in both print and online mediums. The published articles are subject to anonymous and independent evaluation process, and scientific arbitration.

The journal welcomes original empirical research, theoretical and critical views, informative and short notes, which contribute with studies in socio-economic areas, innovative methodology, and analysis in the fields of statistics, dedicated studies in region, and treatment of dealing with issues of common interest in the field of economy, society and especially integration into the European Union.

We are delighted that you are joining us as readers and we look forward to receiving your feedback on the first issue. Furthermore, we encourage you to contribute in the next issues of our journal by providing results of analyses in the socio-economic issues and reflecting the role of official statistics in supporting decision making processes at all levels.

Lastly, we take this opportunity to thank our collaborators, PERFORM – Performing and Responsive Social Sciences – of the Swiss Agency for Development and Cooperation, implemented by Helvetas Swiss Intercooperation and University of Fribourg for their financial support and expertise, and AGENDA Institute for their assistance in the preparation of this publication.

Kind regards,

PhD. Delina IBRAHIMAJ
General Director, Institute of Statistics

FACTORS DETERMINING THE CHILDREN INVOLVEMENT IN UNPAID WORK THE CASE OF ALBANIA

PRANVERA ELEZI, INSTITUTE OF STATISTICS
pelezi@instat.gov.al

Abstract

This article explores the gender roles formation in the division of unpaid work. Unlike previous research on the use of time, this article has in its focus children aged 10 to 17 years old. In Albanian culture, families typically follow a nuclear model comprised of parents and their children. According to 2011 census data, in Albania one tenth of households are with two or more family nucleuses. The issue of whether the presence of grandparents affects the grandchildren gender role formation as regards the unpaid work is very much debated by giving more emphasis to the intergenerational transmission of gender roles. Using data from the 2010-2011 Time Use Survey (TUS) in Albania, we estimate a logit model, to analyze the influencing factors of children involvement in unpaid work. The findings of this study give evidence that gender role formation in the division of unpaid work within the household begins at very prime ages.
Analysis of children gendered division of unpaid work by household composition reveals that although the presence of co-resident grandparents encourages a more balanced involvement of boys and girls in doing household work, the gender gap remains still far away of being egalitarian and it is very skewed, strengthening the definition of household chores as girls' duty. We find evidence that time devoted to household chores increases for children of both sexes, who have both grandparents co-residing in the same home. For girls, the unpaid workload is even higher compared to girls who have only one of the grandparents living in the same household.

KEY WORDS:

Time use; Gender roles; Unpaid work; Gender gap; Family nucleus; Children

1. INTRODUCTION

The question of gender division of paid and unpaid work between women and men and girls and boys has been widely debated in the social science and gender economics (Bianchi, et al., 2006; Olah and Gähler, 2014, etc). Gender equality is the core of development policies and national and international strategies, putting more emphasis on objective of narrowing gender gaps as defined by target indicators such as the gender pay gap, gender gap in labour force participation and gender gap in education (Alvares and Miles, 2008). Even though statistics confirm for a positive progress in gender gaps narrowing, one cannot say the same thing as regards to the gender roles division in everyday activities at home. Women, in the Albanian society, as well as in many other societies, are considered to be the backbone of the household, maintaining home and taking care about children, elderly living in household, and even more caring about their partner/husband (Elezi, 2017).
The study of time use in Albania during the years 2010-2011, corroborated clearly, for the first time in quantitative term, the gender role division concerning paid and unpaid work. Men are seen as financial household pillars that provide income for their family (in other words the breadwinners), while women have to carry out the household chores and take care about children, even if they are employees. Much research has been done on gender role division of unpaid work, focusing mainly on the analysis of who does what at home (Fernandez and Sevilla-Sanz, 2006). However, there are some interesting and relevant problems to be addressed related to unpaid work division within households composed by two or more family nucleuses.
This article concentrates on the question of de/construction gender gap in unpaid work among children aged 10-17 co-residing with their parents (with either parents or one of them) and with at least one of the grandmother/fathers or both. It aims to address the issue of how strong are the correlations of gender role and the intergenerational transmission by controlling for parents employment status and household composition. The influence of the presence of grandparents in obstructing gender division of unpaid work among grandchildren is the one of the hypothesis of this study. Alvares and Miles (2008) explored the hypothesis of transmission of gender roles focusing on how children reproduce their parents' roles at home. This research differs from other empirical analyses of gender role division of unpaid work in that it relies on children information instead of adult information. It extends the time use gender gap analyses by nuclear households.

For the empirical analysis shown in this article, we use time use survey data. Within each sampled household were interviewed all individuals aged 10 years and above, who also completed two diaries. One diary was filled in on a selected weekday and the other diary on a weekend day. The TUS allow matching parents' data with their children data, which is crucial for analysing the inter-generational relations. The remaining of the article is organised as follow: Section 2 provides a brief literature review in the gender roles transmission within the household. Section 3 presents a description of how time use is gendered at very prime ages. Section 4 provides the model specification, followed by section 5 which contains the empirical data analysis. Last section contains main finding drawn from empirical analysis.

2. LITERATURE INSIGHTS

Albanian society is characterised by strong family ties (Bajraba, 2014). Therefore, the transmission process of beliefs and attitudes from parents to their children becomes easier. Alesina and Guiliano (2007) state that strong family ties imply a stricter division of labour with the male working in the market and the female working at home performing a variety of services, probably including maintaining the family ties strong. Gender roles are shaped early in life through the interactions with their environment.

Using the 2002-2003 Spanish Time Use Survey data, Alvares & Miles (2008) found a strong and significant correlation between fathers' and sons' involvement with domestic tasks. They argue that policies aimed at equalizing parents' gender roles may have positive long-term implications through children. Alvares & Miles (2008) points out that a considerable challenge to estimating the casual effect of parental gender roles on children gender roles, measured through household domestic chores, is the endogeneity of father's behaviour because it causes bias of the estimated effect. They argue that it is even more difficult to isolate the father's impact (vertical transmission) from the societal impact (oblique transmission), because the society may also influence fathers' gender roles.

Cosp and Roman (2014) studied the parent-to-child transmission of gender roles in Spain analysing the effects of parent characteristics on the extent to which gender roles are transmitted. They argue that the division of unpaid work in the household is used in many cases as an indicator of behaviour more or less traditional. Whereas, variables such as mother's occupation, father's participation in unpaid work, parents' education explain much of the variation in

child gender roles, supporting an important effect of child-parent mimicry in gendered behaviour.

Dex (2010) raises the concern whether policy intervention in men's and women's shares of unpaid work is possible or realistic. From a labour market viewpoint, she argues that government policies which facilitate the reconciliation of work and family life often play a key role in female labour force participation. On the other hand she accepts that these policies aim to support both parents, but frequently they inadvertently reinforce the more traditional role of women as caregivers, thereby contributing to persistent gender inequality.

Gimenez-Nadal, Molina and Ortega (2017) studied the relationship between parents' time devoted to housework and the time devoted to housework by their children. They showed positive correlations between parents' and children's housework time, indicating that the more time parents devote to housework, the more time their children devote to housework.

Dotti (2016) studied the extent to which daughters and sons learn how to "do gender" in household chores in Italy. She argues that, although both sons and daughters are more likely to engage in housework if their father does so, the effect of paternal involvement is much stronger for sons than daughters. Moreover, she states that the learning of housework is a gendered process and being such it has important implications for the reproduction of gender inequalities in Italy and possibly elsewhere.

In the descriptive and empirical analyses presented in the upcoming sections, when analyzing the factors influencing children involvement in unpaid work activities, are taken into account all the above mentioned factors. In addition to these theoretical aspects, the analysis explores the role of co-residing grandparents (living in the same household with their nephews/nieces), impact in the deconstruction or construction of gender gap in unpaid work.

3. CHILDREN AND UNPAID WORK DIVISION

Time use survey (TUS) is one of the most preferable sources to study gender inequalities in the division of paid and unpaid work in quantitative terms. In Albania, the first survey on time use was carried out during the period 2010-2011. According to TUS results, time use patterns in Albania demonstrate for huge gender disparities regarding the division of household works and the gainful employment or paid work. Time use patterns, by population age structure, revealed that gender inequalities in time use begin at very young ages. Children aged 10 to 14 spend on an average day 53 minutes doing unpaid work, whereas children aged 15 to 17 spend one hour and 14 minutes doing unpaid work.

Based on the TUS estimates, 49% of children aged 10-17, participate in household chores such as household up keeping, shopping, child care, adult care, gardening, laundering and ironing. Also, there are wide gender differences in the proportions of children participating in cleaning, dishwashing, and laundering and ironing. Comparing boys' and girls' participation rates in household chores it seems

Table 1: Children engaged in unpaid work activities

	Age group								
	Aged 10-14			Aged 15-17			Aged 10-17		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
Average hours and minutes per day per person spent on unpaid work									
All	00:23	01:27	00:53	00:18	02:22	01:14	00:21	01:49	01:02
Urban	00:19	00:57	00:39	00:05	01:47	01:00	00:13	01:19	00:48
Rural	00:26	02:02	01:08	00:27	03:19	01:28	00:26	02:29	01:16
Percentage of children participating in unpaid work									
All	30%	67%	48%	24%	81%	50%	27%	73%	49%
Urban	28%	56%	43%	15%	77%	49%	23%	66%	45%
Rural	31%	80%	53%	30%	88%	51%	31%	83%	52%

Source: Albania Time Use Survey 2010-11, Author's calculations, weighted data

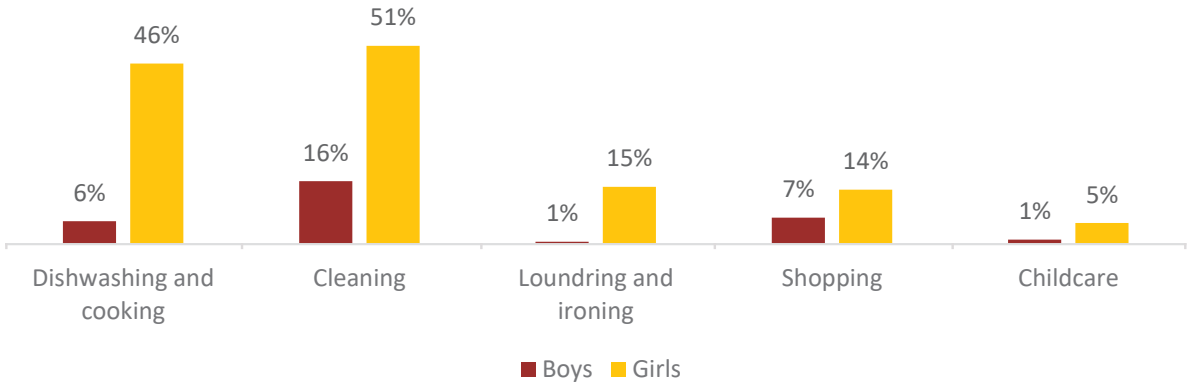
quite evident from the Figure 1 that in Albania these activities are considered as girls' duty. In rural areas the proportion of boys aged 15-17 years doing unpaid work is 30 % while the proportion of urban boys of the same age is 15 %. The level of participation in doing unpaid work is higher among girls aged 10 to 17 (73 %) and for rural girls it becomes even higher (83 %).

The presence of grandparents living in the same household has a controversial impact on the children involvement in unpaid work / domestic chores. According to 2011 census data, in Albania, one tenth of households are with two or more family nucleuses. From a traditional point of view, grandparents are more affectionate to their grandchildren. In Albania it is common to hear from the grandparents the saying "he/she is my honey's honey", phrase that refers to their grandchildren. Grandparents tend to transmit social norms and culture to their grandchildren and feel it like a duty to take care about them and their education.

There are solid preconditions of a plausible effect of grandparents on deconstruction or construction of gender gap in unpaid work of children. Therefore it is important to study the grandparents' role on the nephews/nieces engagement in unpaid work.

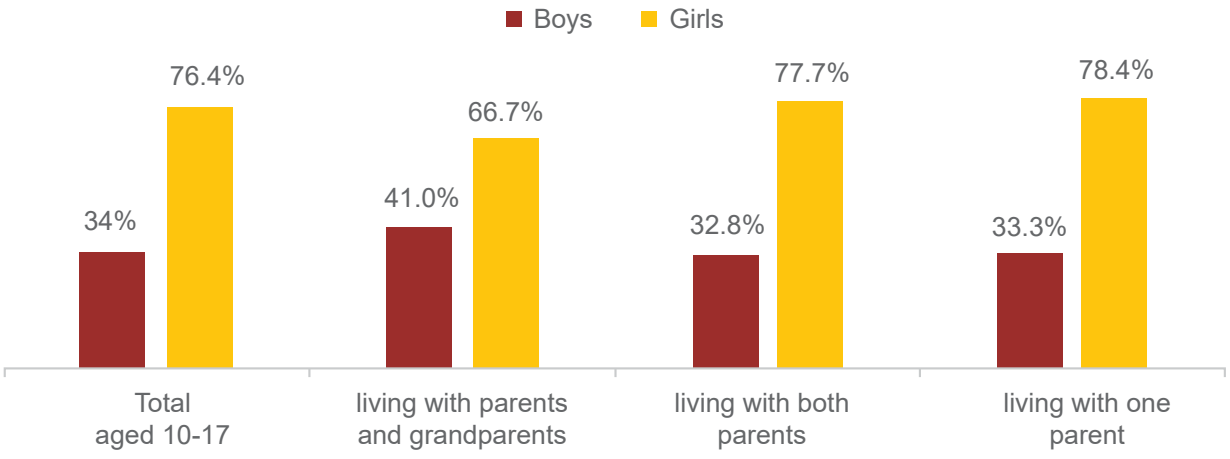
Figure 2 outlines the proportions of children, parents and grandparents undertaking any kind of unpaid work activity by sex and family composition. From this figure it can be seen that boys residing in households with either one or both grandparent have a higher participation rate in doing household chores compared to boys residing in one nuclear household. The difference in percentage points is 8.8. Girls' involvement in domestic tasks is much higher compared to boys in all household types, but, as follows from the figure shown below, the proportion of girls undertaking unpaid work is lower in three-generational households, and even lower compared to girls that live with parents (with a difference of 11 pp.). An important implication of these findings is that grandparents not only engage

Figure 1: Percentage of children aged 10-17 engaged in unpaid work by type of activity and sex



Source: Albania Time Use Survey 2010-11, Author's calculations, weighted data

Figure 2: Proportions of children undertaking unpaid work activities by sex and living status



Source: Albania Time Use Survey 2010-11, Author's calculations, weighted data

themselves in undertaking household chores, but play an important role for stimulating gender equity in unpaid work division within the household.
The average time spent on any unpaid work activity depends on the proportion of people who engage in that activity and the amount of time those individuals spend on the activities. Table 2 summarizes the average time spent on an average day in unpaid work activities by sex and type of household. Gender differences remain significantly high among all generations, even though not in the same amount.

The presence of grandparents in the two or more family nuclei induces a higher participation of boys in household works but still the division of unpaid work in Albania is far way of an egalitarian one (Figure 3).

The participation of fathers in household work is also lower in households with two or more nuclei compared to fathers in nuclear households.

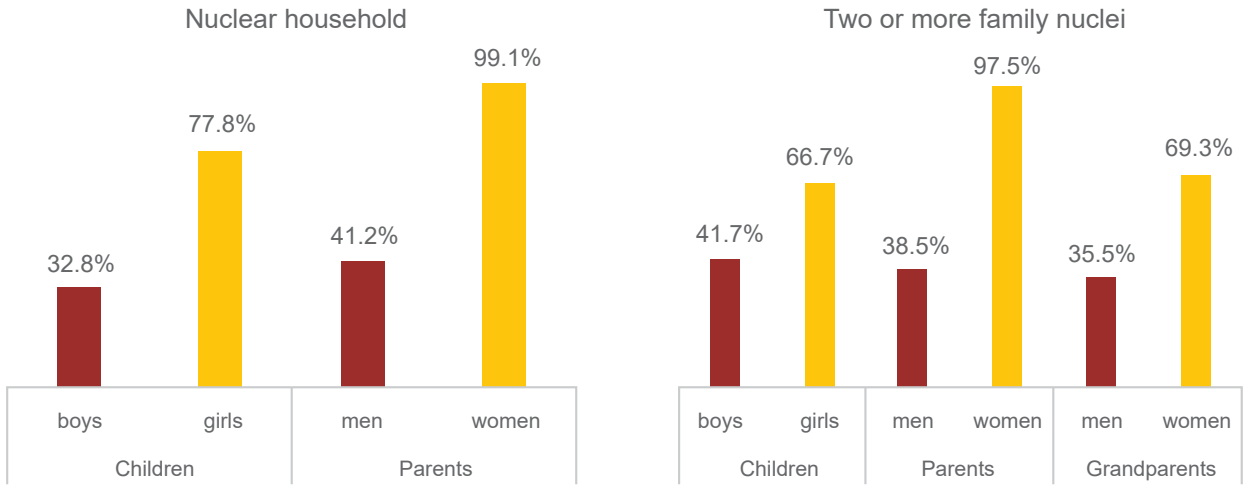
The same holds for girls, who are being less involved in unpaid work if co-residing with grandparents. The lower participation in unpaid work of both fathers and girls in households with two or more nuclei is affected by the involvement of grandparents in doing unpaid work.
The analysis presented hitherto quantifies the time spent and percentage of unpaid work doers. In the next section the analysis is extended by putting emphasise to the potential factors that influence children involvement in unpaid work.

Table 2: Average time spent on an average day in unpaid work by sex and household composition

	Sex	***Nuclear household***	***Two or more family nuclei***	***Total***
		in hours and minutes		
Children	boys	00:31	00:29	00:31
	girls	02:27	02:31	02:28
Parents	men	00:51	00:28	00:47
	women	06:44	06:26	06:41
Grandparents	men	…	00:44	00:44
	women	…	02:51	02:51
Total	men	00:41	00:31	00:39
	women	05:02	04:10	04:50

Source: Albania Time Use Survey 2010-11, Author's calculations, weighted data

Figure 3: Proportions undertaking unpaid work activity by sex and family composition



Source: Albania Time Use Survey 2010-11, Author's calculations

4. MODEL SPECIFICATION

The empirical analysis is based on 2010-11 Albanian Time Use Survey (ATUS). The 2010-11 ATUS is based on a representative probability sample of 2,250 households. The survey covers the Albanian population 10 years and older. Each household member aged 10 years and over was asked to fill in the individual questionnaire and two time diaries. The information on time use was collected by means of a fixed interval (10 minutes) time diary in which the respondents themselves recorded their time use for two randomly designated diary days.

For the analysis presented in this article we use data for children aged 10-17, their parents as well grandparents. The household composition is grouped in two broad categories: (1) nuclear household that is defined as a household consisting entirely of a single family nucleus (both parents with child/children or single mother/father with child/children); and (2) two or more nuclei related to each other (children living with parents, either both or one of them, and with either both grandparents or one of them). These restrictions yield a sample of 954 children, 1,200 parents and 95 grandparents. Unpaid work or household and family care activities throughout this paper refer to the main activity and include food preparation, dish washing, household upkeep, gardening, shopping, childcare, adult care and other related activities.

Included in analysis are some variables that might influence the gender role division regarding to household chores. The literature on gender division of housework demonstrates that the activity status of parents influences the engagement of children in unpaid work activities and has a significant impact on the gender roles formation. In this article the activity status of parents is categorised in four broad categories: both parents are employed; both parents are not in employment; father is employed, mother not in employment; and mother is employed, father not in employment.

Another instrumental variable used in analysis relates to the number of siblings. This variable is grouped in two categories taking value 1 if the child has older sibling/s above the age ten, and value 0 if the child has younger sibling/s less than 10 years old.

We estimate a logit model to explore what determines the gender role division of unpaid work among children. For given vector of p independent variables, the logit of the multiple logistic regression is given by the equation:

$$g(x) = \beta_0 + \beta_1 x_1 + \cdots + \sum_{j=1}^{k_{j-1}} \beta_{jl} D_{jl} + \beta_p x_p$$

the logistic regression model is:

$\Pr(child_{uw} = 1) = \pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}}$, where

$$\begin{aligned} g(x) = \beta_0 + \beta_1 Sex + \beta_2 AGE + \beta_3 siblings \\ + \beta_4 Father_{uw} + \beta_5 GrandFath_{uw} \\ + \beta_6 GrandMoth_{uw} + \beta_7 Fath_{empl} \\ + \beta_8 Parents_{NE} + \beta_9 Moth_{empl} \\ + \beta_{10} nucleus_{Both\ parents} + \beta_{11} nucleus_{one\ parent} \\ + \beta_{12} settlement + \beta_{13} day \end{aligned}$$

Table 3: Description of variables included in the logit model

Variables	Description	Coding
***Child*$_{uw}$ (outcome variable)**	Child participation in unpaid work activities	0 - not participates 1 - participates
Sex	Child gender	0 - male 1 - female
Age	Age of respondent	
Siblings	Presence of siblings	0 – Has younger sibling/s less than 10 years old 1 – has older sibling/s above the age ten
***Father*$_{uw}$**	Father participation in unpaid work activities	0 - does not participate 1 - participates
***GrandFath*$_{uw}$**	Grandfather participation in unpaid work activities	0 - does not participate 1 - participates
***GranfMoth*$_{uw}$**	Grandmother participation in unpaid work activities	0 - does not participate 1 - participates
***Father*$_{empl}$**	Father is in employment, while mother is not in employment	1 - Father employed, mother not in employment 0 - Otherwise
***Parents*$_{NE}$**	Both parents are not in employment	1 - Both parents are not in employment 0 - Otherwise
***Mother*$_{empl}$**	Mother is in employment, while father is not in employment	1 – Mother employed, father not in employment 0 - Otherwise
***nucleus*$_{Both\ parents}$**	A single family nucleus (children live with both parents)	1 – live with both parents (single family nucleus) 0 - otherwise
***nucleus*$_{one\ parent}$**	A single family nucleus (children live with one of parents)	1 – live with one parent (single family nucleus) 0 – otherwise (two or more family nuclei, or in a single family nucleus with both parents
settlement	Settlement area	1 – urban area 2 – rural area
day	Type of day	1 – weekday 2 – weekend day

5. EMPIRICAL RESULTS

Table 4 shows the results of the estimated odds ratios of the logit model that explains the probability of children involvement in any kind of unpaid work. The descriptive analysis showed the gender gap in children involvement in unpaid work and this is confirmed by the logit model estimates which tell us that girls involve significantly in household work. The probability that a girl participates in doing household chores is 7.4 times higher than the one of a boy. The division of unpaid work activities is strongly gendered and skewed.
Having siblings is considered a significant factor influencing involvement of children in unpaid work (Solaz and Wolf, 2015). Alvares and Miles, (2008) argue that the presence of siblings could also affect to the transmission of gender roles. The results of the estimated logit model indicate that the likelihood of involvement in household chores is lower for a child who has older siblings of age ten and above than of a child who has younger siblings aged ten years or less. Usually older siblings take care about their younger siblings.

Most of empirical studies indicate a statistically significant impact of participation of fathers in housework activities (Gimenez-Nadal, Molina and Zhu, 2014). Estimated results show that in the participation of fathers in doing unpaid work has a

Table 4: Estimated odds ratios for children participation in unpaid work activities

Variables	Odd ratios	Standard error (Bootstrap)	P > z
Sex **Ref. categ: Boys**	7.382*	0.926	0.000
Age	1.116*	0.030	0.000
Siblings **Ref. categ: Has younger sibling/s less than 10 years old**	0.644*	0.066	0.000
Father$_{uw}$	1.420*	0.168	0.003
GrandFath$_{uw}$	2.175	0.982	0.085
GranfMoth$_{uw}$	0.792	0.282	0.512
Father$_{empl}$ **Ref. categ: Both parents in employment**	0.725*	0.089	0.009
Parents$_{NE}$ **Ref. categ: Both parents in employment**	0.805	0.153	0.253
Mother$_{empl}$ **Ref. categ: Both parents in employment**	1.083	0.273	0.753
nucleus$_{Both\ parents}$ **Ref. categ: two or more family nuclei**	1.027	0.251	0.914
nucleus$_{one\ parent}$ **Ref. categ: two or more family nuclei**	1.234	0.414	0.531
Settlement **Ref. categ: Urban area**	1.913*	0.187	0.000
Day **Ref. categ: weekday**	1.688*	0.176	0.000
Constant	0.070	0.031	0.000

Number of obs = 1764
Wald chi2(12) = 492.58, Log likelihood = -991.315
Prob > chi2 = 0.0000
Pseudo R2 = 0.1816
*p= 0.05

statistically significant positive effect on children involvement in doing household work. Thus, the likelihood of participation in unpaid work is 1.4 timed higher for a child whose father does any unpaid work at home. This finding supports the hypothesis on parental role model and intergenerational transmission (Sevilla, Gimenez-Nadal, and Fernandes, 2010), where the father's participation in unpaid work affects the transmission of gender roles.

Conforming to the evidence provided in section 3, age has the expected positive effect on children participation in unpaid work and is statistically significant. Children are more engaged in unpaid work activities as they grow up. Considering the economic activity status of parents, the likelihood of children participation in unpaid work decreases for children whose fathers are in employment and mothers are not.

The results provide evidence on the insignificant relationship between grandparents' participation in unpaid work and children participation. The possible limitations of this finding might be related to the observed number of the households with two or more nuclei, and the fact that there is a higher proportion of single family nucleus.

The estimates in Table 4 reveal that the likelihood of participation in any unpaid work is twice times higher for a child of the rural area compared to a child of the urban area. The likelihood that children do engage in household chores is higher in weekend days.

6. CONCLUSIONS

This article studied the determinants of gendered division of unpaid work among children in Albania, using TUS data and estimating a logit model.
Gender inequalities in unpaid work division begin at very young ages, while the empirical analysis shows that the presence of grandparents in gender role formation related to unpaid work activities encourages a higher participation of boys in household works. But despite this, the division of unpaid work in Albania is far way of an egalitarian one.
Comparing boys' and girls' participation rates in household chores it is clearly proved that in Albania these activities are considered as girls' duty.
The division of unpaid work activities is strongly gendered and skewed. The probability that a girl participates in doing household chores is 7.4 times higher than the one of a boy.
The participation of fathers in doing unpaid work has a statistically significant positive effect on children involvement in doing household work. Thus, the likelihood of participation in unpaid work is 1.4 timed higher for a child whose father does any unpaid work at home.
The settlement area is another factor that has a statistically significant positive impact on children involvement in unpaid work activities. Thus, the likelihood of engaging in household chores is higher for a child who lives in rural area compared to a child living in urban area.
Type of day determines children participation in unpaid work, with a higher likelihood if is weekend day.

BIBLIOGRAPHY

Alesina, Alberto & Giuliano, Paola, 2007. "The Power of the Family," IZA Discussion Papers 2750, Institute for the Study of Labor (IZA).

Almudena Sevilla-Sanz, Jose Ignacio Gimenez-Nadal & Cristina Fernández (2010) "Gender Roles and the Division of Unpaid Work in Spanish Households", Feminist Economics, 16:4, 137-184
DOI: 10.1080/13545701.2010.531197

Alvarez, B. and D. Miles (2008) "Intergenerational transmission of gender roles: An empirical analysis through housework".

Bajraba, K, (2014), "Albania, Perspectives and Perceptions of the Mediterranean", The Anna Lindh Report 2014

Bianchi, S. M. (2006). Mothers and daughters "do," fathers "don't do" family: gender and generational bonds. Journal of Marriage and Family 68 (4), 812-816.

Cosp,M.A and Roman, J.G. (2014), "Intergenerational transmission of gender roles in the household", Revista Catalana de Sociologia, (29), 35-47.

Dex, S. (2010), "Can State Policies Produce Equality in Housework?", in Treas, J. and S. Drobnic (eds.), Dividing the Domestic, Stanford University Press, Stanford.

Dotti Sani, G. M. (2016), "Undoing Gender in Housework? Participation in Domestic Chores by Italian Fathers and Children of Different Ages" (Sex Roles (2016) 74: 411. https://doi.org/10.1007/s11199-016-0585-2)

Fernandez, C. and A.Sevilla-Sanz (2006) "Social norms and household time allocation", Economics, Series Working Papers 291, University of Oxford, Department of Economics.
INSTAT (2011) Albania Time Use Survey 2010-11

Elezi, P. (2017) "Unpaid work in Albania and the inheritance of gender roles", working paper presented in the Conference of European Statisticians on "Measuring time use and valuing unpaid work", Belgrade 2017.

Gimenez-Nadal, J.I., J.A. Molina and R. Ortega (2017). "Like my parents at home? Gender differences in childrens' housework in Germany and Spain" Empirical Economics 52(4): 1143-1179.

Gimenez-Nadal, J.I., J.A. Molina and Y. Zhu (2014). "Intergenerational Mobility of Housework Time in the United Kingdom". Review of Economics of the Household (DOI: 10.1007/s11150-017-9374-0).

OECD (2012) Gender Equality in Education, Employment and Entrepreneurship: Final report, OECD Publishing, Paris.

Oláh, L. and Gähler, M. 2014. "Gender Equality Perceptions, Division of Paid and Unpaid Work, and Partnership Dissolution in Sweden". Social Forces, 93(2), 571-594.

Sevilla, A., J.I. Gimenez-Nadal and C. Fernandez (2010). "Gender Roles and the Division of Unpaid Work in Spanish Households," Feminist Economics 16(4): 137-184.

STATISTICAL MATCHING FOR DATA INTEGRATION WITH DIFFERENT DATA SOURCES

EVA XHAVA, INSTITUTE OF STATISTICS
exhava@instat.gov.al

Abstract

During the process of creating databases, which are frequently used by analysts and statisticians, several data files are combined by statistical matching techniques to enrich the host data file. This process requires the conditional independence assumption (CIA) which could lead to serious bias in the resulting joint relationships among variables. In this article, methods of statistical matching are considered. Results are based in real data from the file of Value Added Tax (VAT) and the file of Structured Trade Survey (STS), and for their implementation is used the language of programming R. Based on a case study, analysis of pre-requisites and the results of statistical matching methods are performed, and the gains from using auxiliary information are mentioned. Some conclusions about the methods of statistical matching performed are achieved. Specifically, it was confirmed that the CIA could be a serious limitation which could be overcome by the use of appropriate auxiliary information. Hot deck methods were found to be preferable to other methods performed in this case study. This article was motivated by the need to analyze together turnover data collected by different data sources, and in this case from the file of VAT and STS.

KEY WORDS:

Statistical matching; Auxiliary information; Conditional independence assumption

1. INTRODUCTION

Statistical matching is a model-based approach for providing joint information on variables and indicators collected through multiple sources[1] (surveys drawn from the same population). The potential benefits of this approach lie in the possibility to enhance the complementary use and analytical potential of existing data sources. Hence, statistical matching can be a tool to increase the efficiency of use given the current data collections. Often the aim of a matching exercise is to enlarge the information scope, but matching techniques have also been used for alignment of estimates observed in multiple surveys and for improving the precision of these estimates by integration with larger surveys. Two main approaches can be delineated in terms of outputs that can be obtained through matching:

a) The macro approach refers to the identification of any structure that describes relationships among the variables not jointly observed of the data sets, such as joint distributions, marginal distributions or correlation matrices (D'Orazio, 2006).

b) The micro approach refers to the creation of a complete micro-data file where data on all the variables is available for every unit. This is achieved by means of the generation of a new data set from two data sets that are based on an informative set of common variables between two 'synthetic micro records'. In practice, matching procedures can be regarded as an imputation problem of the target variables from a donor to a recipient survey. Y, Z are collected through two different samples drawn from the same population; X variables are collected in both samples and they are correlated with both Y and Z. The relation between these common variables with the specific variables observed only in one of the data sets - the donor data set will be explored and used to impute to the units of the other data set - the recipient data set - the variables not directly observed. Thus, a synthetic dataset is generated with complete information on X, Y and Z.

In particular, measures of association between Y and Z conditional on X cannot be estimated and they are usually assumed to be 0. This is the so called conditional independence assumption (CIA), a reference point for assessing the quality of estimates based on matching.
When this condition holds, matching algorithms will produce accurate estimates that reflect the true joint distribution of variables that were collected in multiple sources. It will give the same results as a perfect linkage procedure. Unfortunately, this assumption rarely holds in practice and it cannot be tested from the data sets. In case the conditional independence does not hold, and no additional information is available, the model will have identification problems and the artificial datasets produced may lead to incorrect inferences.

Another approach for tackling the conditional independence assumption is the use of auxiliary information. Auxiliary information usually comes in one of the following possible types:
a) Auxiliary parametric information, obtained from "hook"[2] variables.
b) A third data set (C) or an overlap of the two samples (A, B) that provides complete information on (X, Y, Z).

In a macro-matching parametric approach the auxiliary information, generally collected from hook variables, or through previous samples, archives or collection of data, can be particularly useful. Hook variables can contribute to significantly increasing the explanatory power of the common variables and therefore decrease the degree of uncertainty, and can eventually eliminate it completely in some cases. Auxiliary datasets can also be of use in the macro matching approach. The likelihood function can be split into two factors, and the data files A, B and C can be merged into one file. The final report of the ESSnet[3] on Data Integration identifies three main methodologies that focus on the use of auxiliary datasets with complete information:
• Singh et al (1993) proposes a two-step procedure for the use of auxiliary dataset in the context of "hot deck"[4] methods. First, a live[5] value of the variable Z from the data set C is imputed to each unit in data set A using one of the hot deck procedures. Secondly, for each record in A, a final live value from B will be imputed: the one corresponding to the nearest neighbour in B with a distance calculated on the previously determined intermediate value.
• Another methodology for the use of auxiliary information which takes into account complex sample designs is provided by Renssen (1998). Renssen identifies two approaches for providing estimates from the joint dataset, mainly focused on the adjustment of weights:
a) The 'calibration approach' that is obtained under

[1] Donald Rubin

[2] A "hook" variable is a variable in which can be saved a function or several functions which can be used in a special case from an existent program.
[3] European Statistical System
[4] The "hot deck" package contains all the necessary functions to do the hot deck imputations in a set of input data with missed observations using either the best cell method or the probabilistic method.
[5] The live value corresponds to the nearest neighbour

the incomplete two way stratification. This approach consists in calibrating the weights in the complete file (C) constraining them to reproduce in C the marginal distributions of Y and Z estimated from files to be matched.
b) A 'matching approach' where a more complex estimate of P (Y, Z) can be obtained under the synthetic two way stratification. Roughly speaking it consists in adjusting the estimates computed under the conditional independence assumption using residuals computed in C between predicted and observed values for Y and Z respectively.
• The third approach was proposed by Rubin (1986) and consists in appending the two data sources A and B. In the case of an overlap of samples, difficulties in estimating the concatenated weights can limit the applicability of this approach.

There is an important need to analyze together the data of turnover collected from different sources. There are two data files that have these data: Value Added Tax File (VAT) and Structured Trade Survey of Economic Enterprises file (STS). On the one hand, VAT file is an important source for the collection of turnover. On the other hand, STS collects a large range of variables relevant for economic analysis, and one of them number of employees.
This article aims to test the use of alternative model based techniques to integrate turnover information from VAT into STS file.

The objectives are:

Objective 1: Analysis of the coherence between turnover statistics based on currently collected STS turnover and VAT. This comparative overview of STS coherence with VAT shall provide important insights on the quality of the information collected in STS.
Objective 2: Assess the quality of turnover obtained through statistical matching in combination with variables collected in STS.

In the section 2 of the study are presented the main implementation steps of matching, highlighting the main results in relation to the two objectives. Meanwhile section 3 summarizes the main conclusions and recommendations for the application of statistical matching techniques in this article.

The question of the research is: Is there coherence between the statistics of turnover collected from VAT and STS?

In order to accomplish the objectives of the research, there should be tested the hypothesis on the absence or not of coherence between statistics of turnover collected from STS and VAT.

2. STATISTICAL MATCHING: METHODOLOGY AND RESULTS

The two data sources — VAT as donor and STS as recipient — share a set of common variables consistent in terms of definitions, classifications, marginal and joint distributions, and reference period. The two data sources have the same target population - employees.

2.1 METHODOLOGY

2.1.1 Comparison of distributions for common variables

On the basis of the selected target population (employees), the consistency of the marginal distributions of common variables is analyzed. The Hellinger distance metric (HD)[6] has been applied as a yardstick of similarity of distributions for the common variable used in the matching process. Below are presented the values of coefficients that compare two distributions of the common variable, NACE (the classification of activities according to the Nomenclature of Economic Activities, NACE Rev. 2).

```
$meas
tvd    overlap   Bhatt    Hell
 0         1       1       0
```

Dissimilarity index: The dissimilarity index is defined as the total variation distance (tvd) between the marginal distributions, and ranges from 0 (completely similar) to 1 (completely dissimilar). This index represents the fraction of records that are causing differences between the compared distributions. The smaller the dissimilarity index is, the more coherent the marginal distributions of the response variable in the donor and the integrated datasets are. Agresti suggests that as long as the dissimilarity rate is less than or equal to 6% (tvd ⩽ 0.06), the compared marginal distributions could be considered consistent.

Overlap: The overlap is the opposite of the dissimilarity index (sum of overlap and tvd is 1). Its value ranges from 0 (completely dissimilar) to 1 (completely similar). The higher the overlap is, the more coherent the compared marginal distributions are.

[6] In probability and statistics, Hellinger distance is used to measure the similarity between two probabilistic distributions.

[7] Hot deck is an imputation method that deals with data that are missed in which every missed value is substituted with an observed value in a survey from a similar entity.

To clarify, overlap and tvd are complimentary to each other and their sum is equal to 1. Analogously to Agresti's distributions' consistency suggestion (tvd ⩽ 0.06), it can be concluded that an overlap ⩾ 0.94 indicates that the compared distributions can be considered as consistent.
Hellinger's Distance: The Hellinger's distance is a dissimilarity index representing the distance between the two marginal distributions, which is non-negative, symmetric, and lies between 0 and √2. Hellinger's distance (Hd) is mathematically related to tvd by the following equation:

$$Hd^2 \leqslant tvd \leqslant Hd\sqrt{2}$$

Considering this equation and given that tvd ⩽ 0.06, one can derive that Hd ⩽ 0.042. In literature, when Hellinger's distance ⩽ 0.05 the two distributions are considered consistent.
Bhattacharyya Coefficient: The Bhattacharyya coefficient (Bhatt) is a measure of similarity between two distributions, and ranges from 0 to 1. This coefficient could be used to estimate the relative closeness of two distributions. The higher the value of the Bhatt coefficient is, the more similar the distributions are. The Bhatt coefficient can be mathematically related to the Hellinger's Distance through the following equation:

$$Hd = \sqrt{1 - bhatt}$$

Taking into account the limits of an acceptable Hellinger's distance (⩽ 0.05), the Bhatt coefficient would be acceptable if Bhatt ⩾ 0.9975.
In order to quantify the similarity between probability distributions of donor and recipient data, the Hellinger distance has been used, which lies between 0 and 1. Value 0 indicates a perfect similarity between two probabilistic distributions, whereas a value of 1 indicates a total discrepancy. Calibration techniques applied explained a perfect similarity for the common variable NACE, considering that Hellinger metric distance is equal to 0.

2.1.2 Analysis of the explanatory power for common variables

The choice of the matching variables is a crucial point in statistical matching. The conditional independence assumption is the reference point. The fulfilment of this condition guarantees that the joint distribution of matched variables Y and Z will be the same as the one obtained from a perfect linkage procedure.
The set of common variables is made by NACE code (Classification of Activities according to the Nomenclature of Economic Activities, NACE Rev. 2).

2.1.3 Matching methods

Often the goal is to obtain a complete synthetic micro data file through effective imputation of values to the unobserved variables.
The study is based on the data of the fourth quarter of the year 2016 and are tested several imputation methods like mixed methods for performing statistical matching and "hot deck "[7].
Mixed.mtc[8] function implements some mixed methods to perform statistical matching between two data sources that are showed below:

1. In the case of the estimation method under CIA (rho_YZ|X=0) where there are only parameter estimates (micro=FALSE), the estimated correlation matrix is as follows:

Table 1: The matrix of coefficients of correlations estimated under CIA

	NACE	STS120	Turnover
NACE	1.00000000		
STS120	0.04865750	1.000000000	
Turnover	0.04885995	0.002377403	1.00000000

In the table above STS120 refers to turnover in Structured Trade Survey of Economic Enterprises while Turnover refers to turnover in the file of VAT.

2. In the case of the estimation method with partial correlation coefficient (rho_YZ|X=0.5) where there are only parameter estimates (micro=FALSE), the estimated correlation matrix is as follows:

Table 2: The matrix of partial correlation coefficients where there are only parameters estimates

	NACE	STS120	Turnover
NACE	1.00000000		
STS120	0.04865750	1.000000000	
Turnover	0.04885995	0.501188700	1.00000000

3. In the case of the estimation method with partial correlation coefficient (rho_YZ|X=0.5) where there is an imputation step (micro=TRUE), the estimated correlation matrix is as follows:

Table 3: The matrix of partial correlation coefficients where there is an imputation step

	NACE	STS120	Turnover
NACE	1.00000000		
STS120	0.04865750	1.000000000	
Turnover	0.04885995	0.501188690	1.00000000

[8] This function implements some mixed methods to do statistical matching between two data sources.

4. In the case of Moriarity and Scheuren estimation method under CIA where there are only parameter estimates (micro=FALSE), the estimated correlation matrix is as follows:

Table 4: The matrix of partial correlation coefficients where there are only parameters estimates

	NACE	STS120	Turnover
NACE	1.00000000		
STS120	0.04869503	1.000000000	
Turnover	0.04889764	0.002377403	1.00000000

5. In the case of Moriarity and Scheuren estimation method with correlation coefficient equal with -0.15 (rho_YZ=-0.15), the estimated correlation matrix is as follows:

Table 5: The matrix of partial correlation coefficients in the case of Moriarity and Scheuren estimation method

	NACE	STS120	Turnover
NACE	1.00000000		
STS120	0.04869503	1.000000000	
Turnover	0.04889764	-0.150000000	1.00000000

NND.hotdeck function implements the distance hot deck method to match the records of two data sources that share the same variables. This function finds the closest donors computing Euclidean distance on NACE. It creates the synthetic data set filling STS with the turnover of VAT.

Because imputation approaches have usually limited ability to recreate individual level values, results are assessed in terms of preservation of important data distribution aspects and multivariable relationships (Rubin, 1996).

Therefore, to assess the robustness of different methods applied, it is compared the extent to which the observed distributions in the donor (VAT) are preserved in the recipient (STS) files. Hellinger distances are used again to measure the level of similarity of the joint distributions of turnover with key variables.

In a parametric framework, the assumption of conditional independence ensures that data are sufficient to estimate the parameters of the model.

2.2 RESULTS

The quality assessment in the context of matching needs a process approach. Each of the steps (the quality and the coherence of data sources, modelling techniques, matching/imputation algorithms) has a large impact on the quality of results.

When assessing results based on matched datasets, it should be considered the final aim of the analysis and their interpretation according to objectives of the study. Thus, results have to be interpreted in relation to the two-fold objective of the exercise.

Three main criteria were considered throughout the analysis:

1) The consistency of joint distributions (of turnover with matching variables) among
VAT observed, STS imputed and STS observed. The steps are as follow:
a) The comparison between VAT observed and STS observed helps for checking the coherence of common variables. In this particular case, it also helps to assess the quality of turnover information collected in STS with VAT as a benchmark taking into account objective 1.
b) The comparison between VAT observed and STS imputed serves as a quality criterion of matching referencing to objective 2.
c) The comparison between STS observed and STS imputed helps to compare how matching performs, in comparison with collected information in STS.

2) The consistency of different parameters such as totals, means, etc.

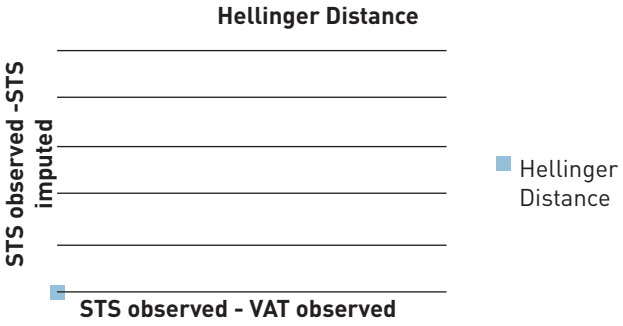
3) Test the CIA for specific target variables: turnover.

The objectives of the study are as follow:

OBJECTIVE 1 - Assess the quality of turnover information in STS with VAT as benchmark

OBJECTIVE 2 - Assess the quality of STS turnover statistics obtained through matching

Figure 1: Average similarity of joint distributions of turnover of STS



To assess the quality of results obtained through statistical matching we refer to two main criteria:

• Preservation of distributions and main parameters between the donor and the recipient.
• Capture the real joint distributions and correlations for variables not collected together.

After an analysis of the imputed turnover information, in general it is noted that the distributional parameters for the turnover variable as well as its joint distribution with matching variables are usually consistent between the donor (VAT observed) and the recipient (STS imputed). For instance, Figure 2 compares the cut-off points for turnover between VAT observed and STS observed for the fourth quarter of 2016. As it is seen, there are similar results both from matching and data collection.

The major limitation of statistical matching is its reliance on implicit assumptions. When imputed turnover need to be analyzed with additional variable collected solely in VAT, one essential condition for success is the existence of good explanatory variables that mediate the relation between these variables.

3. CONCLUSIONS AND RECOMMENDATIONS

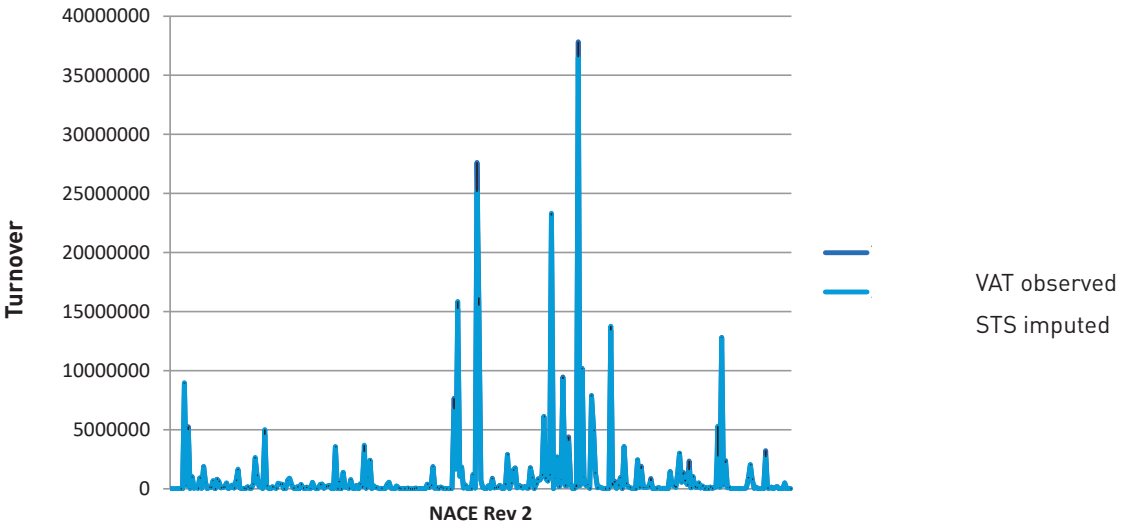
In this article, methods of statistical matching are considered. Results are based in real data from the file of VAT and STS, and for their implementation is used the language of programming R. Based in a case study are performed analysis of pre-requisites and of the results of methods of statistical matching, and are mentioned the profits from using auxiliary information. This article was based in two objectives and the conclusions are as follow:

OBJECTIVE 1: In basis of the analysis, the coherence of turnover of VAT and STS is good.

OBJECTIVE 2: An important factor for the joint analysis and matching of STS and VAT is a better coherence of variables. Differences and misalignment of distributions for the common variables used in the matching algorithm can cause discrepancies for turnover related estimates.

Specific matching methods proved to be more robust. However, results tend to be similar and in general estimates from matching are more sensitive to coherence pre-requisites and variables used in the model than the actual matching method employed.
Results show that, when pre-requisites of coherence are met, matching provides good results for marginal distributions and joint distributions that involve dimensions controlled in the model. However, when model assumptions hold, statistical matching can provide good inferences for specific estimates.
In the case studied, pre-requisites for matching are met. The two data sources - VAT as donor and STS as recipient – are consistent in terms of definitions, classifications, marginal distributions and reference period and the two data sources have the same target population.
Statistical matching should be used because it is a useful method in order to optimize the data sources. It allows using a small measure of sample using a priori stratified analysis with a smaller sample measures, compared with a sample not matched with a posteriori stratified analysis.
Matching avoids a stratified analysis with a lot of strata and indeed, in a not matched case study, while we do logistic regression, we may end up with empty strata.

Figure 2: Graphical illustration of VAT observed and STS imputed for the fourth quarter of 2016



BIBLIOGRAPHY

Andridge R.R., Little R.J.A. (2009) "The Use of Sample Weights in Hot Deck Imputation". Journal of Official Statistics, 25(1), 21–36.

Andridge R.R., Little R.J.A. (2010) "A Review of Hot Deck Imputation for Survey Nonresponse". International Statistical Review, 78, 40–64.

D'Orazio, M. (2014), StatMatch: Statistical Matching (aka data fusion). R package version 1.2.2. http://CRAN.R-project.org/package=StatMatch.

D'Orazio M. (2010) "Statistical matching when dealing with data from complex survey sampling", in Report of WP1. State of the art on statistical methodologies for data integration, ESSnet project on Data Integration, 33–37, http://www.essnet-portal.eu/sites/default/files/131/ESSnetDI_WP1_v1.32.pdf

D'Orazio M., Di Zio M., Scanu M. (2006a) "Statistical matching for categorical data: Displaying uncertainty and using logical constraints". Journal of Official Statistics 22, 137–157.

D'Orazio M., Di Zio M., Scanu M. (2006b) Statistical matching: Theory and practice. Wiley, Chichester

D'Orazio M., Di Zio M., Scanu M. (2008) "The statistical matching workflow", in: Report of WP1: State of the art on statistical methodologies for integration of surveys and administrative data, "ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data", 25–26. http://cenex-isad.istat.it/

D'Orazio M., Di Zio M., Scanu M. (2010) "Old and new approaches in statistical matching when samples are drawn with complex survey designs". Proceedings of the 45th "Riunione Scientifica della Societa' Italiana di Statistica", Padova 16–18 June 2010.

D'Orazio M., Di Zio M., Scanu M. (2012) "Statistical Matching of Data from Complex Sample Surveys". Proceedings of the European Conference on Quality in Official Statistics - Q2012, 29 May–1 June 2012, Athens, Greece.

D'Orazio M., Di Zio M., Scanu, M. (2005) "A comparison among different estimators of regression parameters on statistically matched files trough an extensive simulation study". Contributi Istat, 2005/10

D'Orazio, M. (2014), StatMatch: Statistical Matching (aka data fusion). R package version 1.2.2. http://CRAN.R-project.org/package=StatMatch.

D'Orazio, M., Di Zio, M., Scanu, M. (2006), Statistical Matching: Theory and Practice. John Wiley & Sons, Chichester, ISBN: 0-470-02353-8.

Moriarity C., Scheuren F. (2001) "Statistical matching: a paradigm for assessing the uncertainty in the procedure". Journal of Official Statistics, 17, 407–422.

Moriarity C., Scheuren F. (2003). "A note on Rubin's statistical matching using file concatenation with adjusted weights and multiple imputation", Jour. of Business and Economic Statistics, 21, 65–73.

R¨assler S. (2002) Statistical matching: a frequentist theory, practical applications and alternative Bayesian approaches. Springer Verlag, New York.

Renssen R.H.(1998) "Use of statistical matching techniques in calibration estimation". Survey Methodology 24, 171–183.

Renssen, R. H. (1998), "Use of Statistical Matching Techniques in Calibration Estimation". Survey Methodology, No 24, pp. 171-183.

Rubin D.B. (1986) "Statistical matching using file concatenation with adjusted weights and multiple imputations". Journal of Business and Economic Statistics, 4, 87–94.

Rubin, D.B. (1986) Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations, Journal of Business and Economic Statistics, 4, 87- 95.

Scanu M. (2008) "The practical aspects to be considered for statistical matching". in: Report of WP2: Recommendations on the use of methodologies for the integration of surveys and administrative data, "ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data", 34–35. http://cenex-isad.istat.it/

Singh A.C., Mantel H., Kinack M., Rowe G. (1993) "Statistical matching: use of auxiliary information as an alternative to the conditional independence assumption". Survey Methodology, 19, 59–79.

Singh, A.C., Mantel, H, Kinnack, M and Rowe, G. (1993) Statistical matching: Use of auxiliary information as an alternative to the conditional independence assumption, Survey Methodology, 19, pp 59-79

INVOLVING INNOVATION IN DATA COLLECTION PROCESS IN PRODUCING OFFICIAL STATISTICS

ALMA KONDI, INSTITUTE OF STATISTICS
akondi@instat.gov.al

HELDA CURMA, INSTITUTE OF STATISTICS
hmitre@instat.gov.al

Abstract

Albanian Institute of Statistics (INSTAT) vision is to provide reliable and comparable data, adapting methodologies and adding a list of statistical indicators. Based on this vision the use of technology in different statistical processes has been increased during the last years, starting from data collection, editing, estimation, tabulation and dissemination. INSTAT has made progress in its use of technology (CAPI method, web-based questionnaires, OCR, R, PxWeb).

This article review the situation in 2016 regarding use of innovative technologies, improvements made in comparison with traditional methods of data collection, processing and publishing, and looks forward to the next challenges of INSTAT. The method used in this article is a deep analysis of cost, response rate, data capture and quality of the data in different methods of data collection.

The use of innovation can be made by the statistical offices to reduce cost and response burden, to maximize the quality of the data collected and improves timeliness.

KEY WORDS:

INSTAT; Data Collection; Technology; Statistical Quality

1. INTRODUCTION

One of the most important statistical office's challenges is increasing quality and timeliness in a more cost effective way. The process of increasing high quality data for NSOs' (National Statistical Offices), is embracing the best practices in the collection, reducing cost and response burden. Costs are very important, especially for statistical offices with low budgets. The new technological changes can affect every statistical process. The rapid changes in technology have created new opportunities for improving timeliness. Costs of innovative technologies, such as tablets or other mobile devices are declining, making possible the use on the data collection process through the Computer-Assisted Personal Interviewing (CAPI) method.

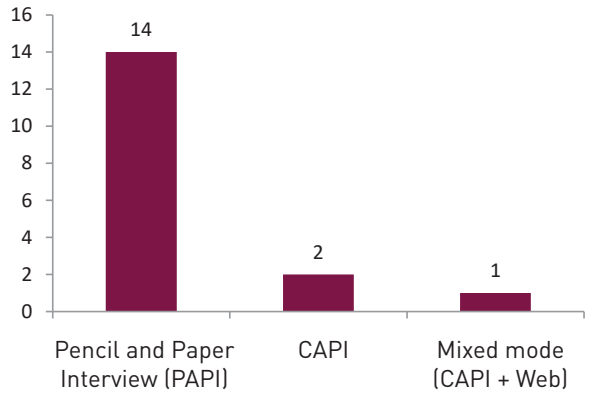
Statistics Canada analyzes in the Methodology Report (Statistics Canada, 2010) the advantages of using technology in statistical production. Technology can monitor and control the quality of the data from data collection, data capture and also in the process of editing and imputation. Also, automated skip patterns can be used to resolve immediate problems, resulting in less follow-up and reduce the burden of the interviewer. It is easier to implement quality control of the interviewing process and generate management reports on the status of the interviews e.g., response rate, number of interviews completed, and length of time per interview, etc. On the other hand, the uses of technology require investment of buying or renew. It also requires computer experts to develop programs for the statistical production.

INSTAT's work is based on the five year Program of Official Statistics, which is the basic document that provides the production of statistical data by the National Statistical System, measuring different statistical process of economic, social and environmental situation in the Republic of Albania. This program works in accordance with the statistical principles provided in the Statistical Law but also follows the general principles of the European Statistics code of Practice.

During 2016, seventeen surveys were conducted by INSTAT. In Figure 1 are shown the number of surveys by the data collection methods for the year 2016.

Traditionally INSTAT has been using Personal Assisted Paper Interview (PAPI) as a data collection method. Nowadays it is still the most used method, as shown in Figure 1; around 82 percent of surveys are using the PAPI method to collect information, due to several factors such as generally high response rate, poor system of addresses, user friendly and often because it is cheaper.

Figure 1: Number of surveys by data collection methods



Source: INSTAT Program for 2016, calculation of the authors

For the first time CAPI was used in 2008-09 Albania Demographic and Health Survey (ADHS). An innovative aspect of the 2008-09 ADHS was the use of Personal Digital Assistants (PDAs) for the data collection, rather than paper questionnaires. The questionnaires were programmed in PDAs using the software package Census and Survey Processing (CSPro). Full survey and data management, range, skip and consistency checking were built into the data capture system. Paper questionnaires were available to interviewers in case of equipment failure. In 2016 the CAPI methods were used by 2 surveys.

The online data collection (WEB) was first used in 2013, as a way to lower costs, avoiding the use of interviewers, and reducing respondent's burden. For this purpose, an open source solution, Limesurvey Server was used. Various forms have been prepared for the collection of information, for example Short Term Statistics Survey, survey of Research Development for Public Sector, etc. This method was used in 2016 in one survey but was also used with CAPI method to provide more information.

2. COMPARISON OF DATA COLLECTION METHODS

This section will describe the different data collection methods, their advantages and disadvantages in different surveys conducted by INSTAT, in various perspectives such as cost, response rate, response burden, quality indicators, etc. These aspects are addressed in this article by analyzing and comparing them in the following surveys:
- Labour Force Survey (LFS)
- Short term Statistics Survey (STS)
- Information Communication and Technology Survey (ICT)

LFS was conducted for the first time in 2007. During 2007-2011 LFS has been carried out by INSTAT on annual basis through direct personal interviews, contacting households at their dwelling. In 2011, for the first time, LFS started using CAPI method. The optimized method (in our case CAPI method), allows us to have more accurate data because logical controls have been directly done in the application of the laptops. These controls reduce non-sampling errors (e.g. filter questions are skipped automatically and not manually, as it happens in paper questionnaires). For example, if a person is younger than 15 years old, the questionnaire regarding employment status is skipped automatically.

STS Survey is a quarterly survey which is addressed to the enterprises. From 2003-2013 the PAPI method has been used to collect the information. Since 2013 efforts has been made to use a mixed method approach: PAPI + online questionnaires. In 2014, 91% of companies, with 10 + employed, have had internet access. The high internet penetration rate, especially among big companies, is the reason why the web approach is being used for this survey. PAPI is applied as the base method, while web forms are being used for the big companies only (approximately 300).

ICT Survey was conducted in 2015 for the first time, using tablets (CAPI) to collect the information. This survey is focused on the availability of information and communication technologies (ICTs), and their use by enterprises.

2.1. COST

There is a growing demand on statistical data from users, not only from policy makers but from other type of users as well. Every statistical agency needs to make decisions between producing statistical data and the costs to do it. Collecting such data is costly, especially in countries with poor system of addresses and infrastructure. Costs affect every statistical process. In this article data collection costs are taken into consideration, as generally this process is the most expensive one. The budget items taken into account to analyze cost per questionnaire (in ALL) are shown in the Table 1.

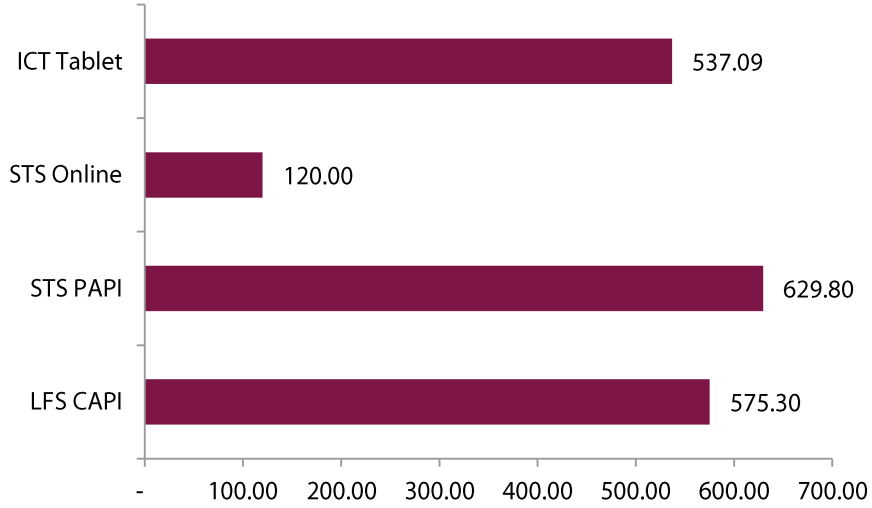
Table 1: Surveys budget items taken into account

	LFS CAPI	STS PAPI	STS Online	ICT Tablet
Printing of questionnaires	NA	Value	NA	NA
Printing of additional materials	Value	Value	NA	Value
Interviewer	Value	Value	NA	Value
Interviewer No	Value	Value	NA	Value
Controller	Value	Value	NA	NA
Consumables	Value	Value	NA	Value
Per diem for interviewers	NA	Value	NA	Value
Data Entry Operator	NA	Value	NA	NA
Internet Cards	Value	NA	NA	Value
Postal Service	NA	NA	Value	NA
Total	Value	Value	Value	Value

Source: Extracted by the authors based on INSTAT information

Figure 2 shows the surveys costs per questionnaire. As it is expected, the online approach by using web forms is the cheapest one. This approach is being used until now only for STS big companies, knowing for sure that they have in place the infrastructure to respond the questionnaire and also the statistical office has the possibility to send them their authentication credentials by post or email. Costs for CAPI using laptops or tablets are almost the same, excluding the initial costs for equipment, which can be very different, depending from the technologies used; although it is decreasing very rapidly. The PAPI method, which is still the most used method at INSTAT has the highest cost, from the other collection method.

Figure 2: Costs per questionnaire according different data collection approaches and surveys (in ALL)



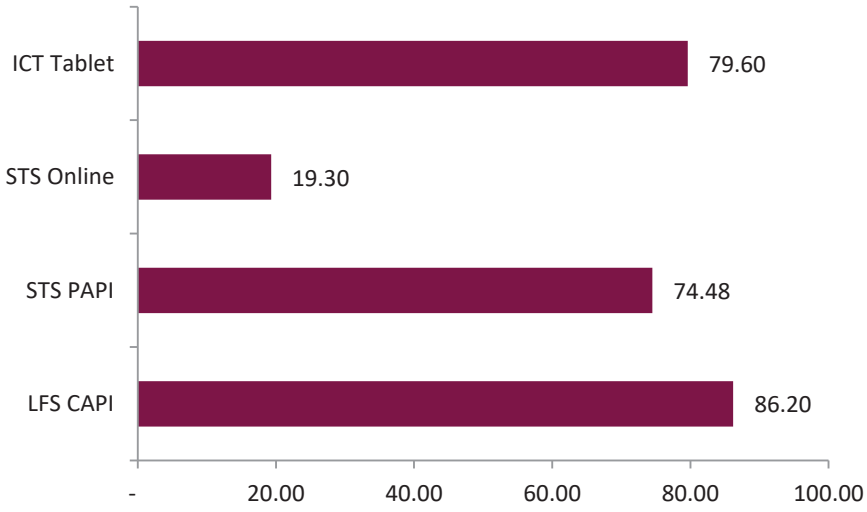
Source: Authors' calculations

2.2. RESPONSE RATE

Non-response is a common feature of sample surveys. A part of the sampled elements cannot be contacted, refuses participation or does not participate in the survey for other reasons. (Fannie Cobben, 2009)

In Albania, to complete voluntary a questionnaire is relatively a new culture. Generally, people cooperate with interviewers to fulfil the questionnaires. Still, it had become difficult to find respondents at home during daytime hours as the lifestyle has changed. The population has become more mobile and people are not at their usual place of residence for extended periods. All these changes can affect the response rate. For this purpose it is calculated the response rate shown in the Figure 3. The response rate is calculated by dividing the number of completed questionnaires by the total number of eligible units in the sample chosen.

Figure 3: Response rates according different data collection approaches and surveys (in %)



Source: Authors' calculations

There is a specific situation in Albania, because the system of addresses has serious weaknesses. This makes it very difficult to mail questionnaires or authentication credentials for online forms. For household surveys, INSTAT is using its own GIS system, designed for Population and Housing Census 2011 (PHC). The response rate for LFS is higher due to this fact. The response rate for LFS is the annual average of 2014, for STS PAPI is the annual average of 2013, for STS online is the rate for fourth quarter 2015, while for ICT tablet is the rate of 2015.

Bradburn (1978) suggested that the definition of respondent burden should include four elements: interview length; required respondent effort; respondent stress; and the frequency of being interviewed. The effort required of respondents could refer to the cognitive challenges of a task. The smallest response rate is for the online approach; 19.3%. Even with the online surveys, observation units can respond when it is appropriate for them, the burden on respondents is higher, as they need to understand, complete, keep notes, and calculate questions.

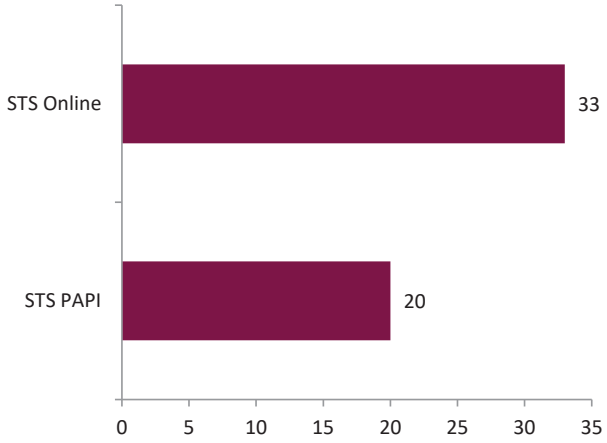
Traditionally, using interviewers for data collection has resulted the most efficient way regarding response rates; approximately 80%. The respondents trust interviewers and collaborate with them better, resulting in higher response rates. Interviewers are equipped with identification badges and have with them a formal paper explaining the scope of the survey and the data confidentiality aspects.

2.3. RESPONSE BURDEN

The measurement of response burden is done by taking into account duration of interview, for the same survey, in different data collection methods. STS survey has been tested in two different methods; PAPI and online. It is used the same questionnaire for both methods. On the online form there have been added additional notes, to explain what the interviewer does in the face to face method. The online form was directed to the same individual as per PAPI method (during PAPI method contact points have been decided previously).

Obviously the online method has a higher burden on the observation unit (in our case companies). This relates mostly to the fact that the respondent needs more time to understand the questions.

Figure 4: Average time spent on PAPI and Online method (in minutes)



Source: Authors' calculations

2.4. DATA CAPTURE

High dynamic developments in Information Technology (IT) and rapid advances in technology changes are affecting the way of work for the statistical offices. INSTAT's everyday challenge is to identify trends and find the most suitable solutions to produce statistics in a more cost effective way. Almost every survey carried out by INSTAT is using the face to face interview. To facilitate the data capturing process, a scanning system was installed in October 2009 to prepare the coming work for scanning censuses and large surveys questionnaires.
INSTAT's achievements regarding data capture are:
- In 2011 scan and capture all data from the Enterprise Census
- In 2011-2012 scan and capture all data from the PHC
- In 2013 scan and capture all data from the Agriculture Census
- From 2010 – till present scan and capture periodically surveys

The Intelligent Character Recognition (ICR) approach reduces time for data capturing process, especially for big statistical activities. For National Statistical Offices (NSO's) there is always a trade between timeliness and data quality. Regarding this, an analysis has been made to estimate the data quality, for Enterprise Census.

When analyzing scanned data, there are several ways to divide the amount of data by questionnaire, page, field or character level. It was chosen the most detailed level, character level, where registered characters, in this case in the transfer file, is compared with characters from questionnaire. Boxes were picked randomly and from these boxes there were questionnaires randomly picket out. All

characters were compared and looked for remaining substitutes, i.e. characters incorrect interpreted or incorrect registered.

Figure 5 shows how the interpreter interprets "6" as "0". If the verifiers have accepted the "0" or change it to e.g. "8", this was a remaining substitute. The fields were divided into three different kinds of fields; handwritten, checkbox and Optical Character Recognition (OCR) (i.e. pre-printed).

Figure 5: Example of a substitute

Source: Enterprise Census, data capturing process

In Table 2 are shown the number of controlled questionnaire in order to analyze the scanned data for the Enterprise Census.

Table 2: Controlled questionnaire

Box Number	Number of questionnaires	Number of checked questionnaires	Share of checked questionnaires
4101	289	14	4.84
4601	252	7	2.78
5101	245	12	4.90
5501	296	8	2.70
6201	162	4	2.47
3902	259	14	5.41
4702	68	8	11.76
Total	**1571**	**67**	**4.26**

Source: Authors' calculations

In the Table 3 are shown the number of substitute's characters in order to analyze the quality of data capture for the Enterprise Census.

Table 3: Result of substitutes

Type of fields	No of characters	No of substitutes	Share of substitutes
OCR numeric	1474	0	-
Handwritten	18610	28	0.15
Checkbox	4110	1	0.02
Total	24194	29	0.12

Source: Authors' calculations

The main purpose of controlling output data was to measure the quality by counting substitutes. Few numbers of substitutes were found as is shown in the Table 3. Handwritten fields represented most of found substitutes, which was not much of a surprise. A typical example was the letter "I" that had become a number "1" instead. OCR fields had no substitutes at all, only one occurrence of substitute for the checkboxes was found. The Table 3 reflects a very good quality of the captured/transferred data, only 0.12 percent of characters were substitutes.

2.5. QUALITY OF THE DATA

Users of survey data should always have at least some basic information about the degree to which survey data were modelled or estimated by imputation. At the end of imputation, it may be useful to produce the following indicators (Statistics Canada, 2010):

- the number of records which were imputed (i.e., the number of recipient records);
- the number of times each field was imputed and by what method;
- the number of records eligible to be used as donors;
- the number of records actually used as donors and the number of recipients each of these donor records imputed;
- a list (or file) indicating which donors were used for each recipient (to trace the sources of unusual imputed records);
- a list of all records for which imputation failed (e.g., because no donor was found).

Eurostat proposes the following standard quality indicators that can be used from the point of view of the producers, for summarizing the quality of the statistics as reported according to the Standard Quality Report (ESTAT/02/Quality/2005/9/Quality Indicators). This set of indicators can be used to measure and follow over time the quality of the data produced in the European Statistical System (ESS)[1]. The standard quality indicators proposed by Eurostat are:

- User satisfaction index
- Rate of available statistics
- Coefficient of variation
- Unit response rate (un-weighted/weighted)
- Item response rate (un-weighted/weighted)
- Imputation rate and ratio
- Over-coverage and misclassification rates
- Geographical under-coverage ratio
- Average size of revisions
- Punctuality of time schedule of effective publication
- Time lag between the end of reference period and the date of first result
- Time lag between the end of reference period and the date of the final results
- Number of publications disseminated and/ or sold
- Number of accesses to databases
- Rate of completeness of metadata information for released statistics
- Length of comparable time-series
- Number of comparable time-series
- Rate of differences in concepts and measurement from European norms
- Asymmetries for statistics mirror flows
- Rate of statistics that satisfies the requirements for the main secondary use

The set of indicators were considered by INSTAT to perform the assessment of the effects of the cleaning procedure (editing and imputation) at aggregate level, and can be grouped into three different kinds (INSTAT Quality Dimensions, 2014):

1. Indicators on the amount of data submitted to the imputation procedure, like Number of Records, Number of Variables, Number of Variables subject to the Imputation procedure, and Number of Total Values.
2. Indicators for the evaluation of the overall effects of the imputation procedure in percentage:
- Imputation rate (I): (Number of Imputed values/ Number of Total values)*100;
- Addition rate (Ia): (Number of Additions/Number of Total values)*100;
- Modification rate (Im): (Number of Modification/ Number of Total values)*100;
- Elimination rate (Ie): (Number of Eliminations/ Number of Total values)*100.
3. Synthetic indicators on the imputation rate by records, like for instance Number of Records with Imputation rate greater than 2% and Number of Records with Imputation rate greater than 5%.
From the indicators shown in Table 4, the LFS has around 0.01% Imputation rate, so CAPI data collection method is a very good method to monitor and control the quality of the data. PAPI, as expected, is the method where imputation and editing have the highest figures, around 4.71% of the total values.

[1] https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality_reports/reference_documents/STANDARD_QUALITY_INDICATORS.pdf

Table 4: Quality assessment indicators at aggregate level

	LFS CAPI	STS PAPI	STS Online	ICT Tablet
Number of Records	12,490	6,803	60	1,473
Number of Variables	164	59	59	93
Number of Total values	2,048,360	401,377	3,540	136,989
Number of Imputed values	**20,484**	**1,890,486**	**3,929**	**56,165**
Number of Additions	-	545,873	2,443	5,480
Number of Eliminations	-	8,028	-	16,439
Number of Modification	20,484	1,336,585	1,487	34,247
% Imputation rate (I)	**0.01**	**4.71**	**1.11**	**0.41**
% Additions rate (Ia)	-	1.36	0.69	0.04
% Elimination rate (Ie)	-	0.02	-	0.12
% Modification rate (Im)	0.01	3.33	0.42	0.25
% Non-Imputation rate	**99.99**	**95.29**	**98.89**	**99.59**
% of records with I greater than 2%	**5**	**115**	**26**	**112**
% of records with I greater than 5%	**5**	**13**	**18**	**9**

Source: Authors' calculations

It is obvious from the above data that the CAPI method benefits are not only on the processing time (as data entry is done during data collection process) but also on the data quality.

3. OTHER STATISTICAL PROCESS, BESIDES DATA COLLECTION

INSTAT is aiming at replacing SAS and SPSS software in favour of R statistical software package which is an open source tool capable of doing most of the same thing as the others. This shift is mainly due to the very expensive license costs. In the meantime, before the process is fully accomplished, there is still a need for use of SAS and SPSS. So the plan is to gradually switch from SAS/SPSS to R.

In September 2012 INSTAT has developed a new well-functioning website with user friendly structure. It means using the Common Nordic Meta Model (CNMM) to allow user's access to statistical data, dynamically. INSTAT's statistical databases have been built with a PX-Web user interface. From January 2016 this tool is free of charge, while PX-Web is free of charge since January 2015. PX-Web is developed and owned by Statistics Sweden and is used to establish dynamic tables[2].

Every previously mentioned effort is done to contribute to the development of a sustainable statistical system in Albania. This system should facilitate decision-making based on relevant and reliable statistical information that meets domestic needs.

[2] http://www.scb.se/sv_/PC-Axis/Programs/PX-Web/

4. CONCLUSIONS

PAPI method has a good response rate; the interviewer has a positive effect on this aspect. INSTAT interviewers are equipped with additional materials (maps, formal papers, etc.) in order to improve the quality of data. Based on the quality assessment, complex questionnaires (skipping rules, rosters), are difficult to be completed.
CAPI method benefits from the positive effect of using interviewers, especially in response rates. INSTAT interviewers are equipped with additional materials (maps, formal papers, etc.). The data entry process is done during the data collection phase and more complex questionnaires can be used. Regarding the quality of the collected data, this method reduces interviewer's errors, as range and consistency rules are applied during the data collection. Equipment used for this method, even though prices are declining, is still costly. Using laptops or other kind of equipment in fieldwork is not very easy and a good training is needed. Interviewers and respondents have reacted positively to the use of laptops and tablets.

The data produced for this analysis support the assumption that the cost of laptops or tablets used for data collection process is divided by different surveys that use this equipment. The online method has the lowest cost and this method has also the lowest response rate. The burden on interviewers is higher on the online method, due to the fact that it requires more time to read and understand the questions.

CAPI and online method allow instant data transmission to headquarters, making possible immediate further processing of data.
Including innovation in the process of collecting data for the production of official statistics should be done by analyzing cost, quality of the data and processing time.

BIBLIOGRAPHY

Bradburn, N. M. "Respondent Burden", Paper presented at the 138th Annual Meetings of the American Statistical Association, San Diego, CA., 1978

ESTAT/02/Quality/2005/9/Quality Indicators,

Fannie Cobben, Statistics Netherlands, Nonresponse in Sample Surveys, 2009

https://circabc.europa.eu/webdav/CircaBC/ESTAT/prodcom/Library/25-Quality/quality_reports/reference_documents/STANDARD_QUALITY_INDICATORS.pdf

INSTAT, Quality Dimensions of the 2011 Population and Housing Census of Albania, May 2014

Statistics Canada (2010) Survey Methods and Practices. Statistics Canada, Catalogue no. 12-587-X, Ottawa. http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.htm

THE HOUSING MARKET IN ALBANIA: HEDONIC REGRESSION

ILIRJANA KRAJA, INSTITUTE OF STATISTICS
ikraja@instat.gov.al

MIRJANA SEJDINI, PHD HIGHER COLLEGES OF TECHNOLOGY, UAE
msejdini@hct.ac.ae

ABDULMENAF SEJDINI, PHD HIGHER COLLEGES OF TECHNOLOGY, UAE
asejdini@hct.ac.ae

Abstract

Housing market is one of the most important sectors in Albania. During the last 20 years there has been a boom of construction in this sector, mainly in big cities. The crisis that swept the world, including Albania, was reflected in this sector as being associated with sensitive decrease on housing sales. A phenomenon seen in the country is that despite the declining on purchasing power, constructors do not decrease their housing prices. The main objective of this article is to apply the Hedonic Regression Method as an alternative valuation method for forecasting housing prices in Tirana. The empirical results show that most of the independent variables on housing features are highly significant and economically consistent with the pricing theory. This article contributes to the literature by applying a standard evaluation methodology on a market with unique features as there are very few studies that employ the hedonic regression on the housing prices in Albania.

KEY WORDS:

Albania; Housing market; House prices; Hedonic Price Method

1. INTRODUCTION

Housing sector/market is a very significant part of national economy and is closely associated with other sectors of the economy that together determine the level of development. Housing market's importance is seen in its causal relationship with other sectors of the economy because a demand for housing represents both an investment and a consumption decision. Housing being as the fulfilment of a basic need influences other needs such as food, clothing, water, sanitation and health. Housing market creates complementary sectors and markets such as construction of water, electrical services, shopping complexes and other community services which in turn generate more employment and income. Likewise, housing as a capital asset, like all other investments, provides and helps in the development of business enterprises, provides stable long term rental income and may act as a credit collateral for business endeavours.

Since housing market is an important sector of the economy and it has very significant impact on other sectors of the economy, the research for modelling the housing prices becomes a very important issue. The main objective of this article is to apply the Hedonic regression (price) Method as an alternative valuation method for housing in Tirana. Next section provides an overview of the Albanian housing market particularly its capital Tirana. Section three, provides literature review explaining the theory behind the HPM and the main empirical studies pertaining to the issue. Section four gives a detailed explanation of the data and the methodology used and provides the main empirical findings. The final section is reserved for the conclusions and provides specific explanations for certain phenomena acquired from the results.

2. OVERVIEW OF THE ALBANIAN HOUSING MARKET

During the last two decades, Albania has experienced a significant boom of construction sector, mainly in housing market. Transition period was accompanied by changes in the demographic, economic and social factors, which in turn have increased demand for new construction. Housing market represents an important element not only for Albanian economy but also from the Albanian culture. They consider the proprietorship of a house as one of the primary and initial needs. In this regard, internal (from rural to urban) and external migration and the change in the attitude of living separated from parents has played a very important role in the shape and trend of housing demand. More specifically, emigration has played a major role in the demand for houses in urban areas, especially Tirana, as a consequence of the improvement in living conditions and higher disposable income of Albanian emigrants and their families at home.

Important issues that require specific attention about the Albanian housing market are:

- The ownership rights not clearly defined between "old" owners deprived from the communist regime and the new ones currently occupying the property
- Legalization issue of the constructed buildings without proper local authorities permission
- Mortgage registration, mortgage lending and mortgage execution process which are particularly related also with the banking practice and legal framework in the country.

Based on the results of Population and Housing Census 2011 (INSTAT, 2012) (INSTAT, 2012), the usual resident population in Albania was 2,821,977 and 25.25% of which are residing in Tirana. The total number of dwellings is 1,012,400, from which, according to classification by type, 99.6% were conventional dwellings and 0.4% of them were non-conventional ones. Tirana had the largest percentage of dwellings (26%) and the largest number of building permits (28% in 2017).

Table 1: Population and housing growth, Census 1989, 2001, 2011

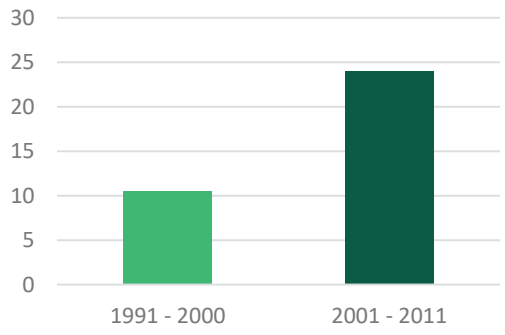
	Census 1989	Census 2001	Census 2011	Census2001/ Census 1989	Census2011/ Census 2001	Census2011/ Census 1989
Residential Population	3,182,417	3,069,275	2,821,977	-3.6	-8.8	-12.8
Buildings	385,769	512,387	598,267	32.8	16.8	55.1

Source: INSTAT

Buildings' features in 2011 in Albania do not vary drastically from 2001 in terms of type of building, number of floors, number of dwellings per building etc. (INSTAT, 2014) (INSTAT, 2014), as shown in Figure 1.

The number of individual buildings still prevails, while apartment buildings constitute only 3.7% in 2011, very close to 3.3% in 2001. The type of buildings and the related period of construction show that the number of apartment buildings constructed during 2001-2011 has more than doubled compared to the previous decade and it accounts for 23.4% of the total stock of that type of building in Albania. Figure 1 shows a boom in construction during 2001-2011compared to 1991-2000.

Figure 1: Number of apartment buildings by period of construction, Census 2011 (in %)



Source: INSTAT: Albania dwelling and living conditions 2014

3. LITERATURE REVIEW

Evaluation of the house prices has attracted a significant interest in economic literature and consequently there is plenty of research work. Hedonic demand theory also known as hedonic regression or as hedonic price method estimates the value of the characteristics of a commodity that indirectly affects its market price. The basic idea of hedonic price method is that it breaks down the house value or its price into its constituent characteristics or properties and then adds up all of the estimated contributory values of each characteristic to create the final value of the house. Simply put, these pricing method measures the weight or the impact of certain feature like location of the house, number of bathrooms, or the presence of other features in the total price of the house.

This theory of pricing method is widely used in consumer and market research. From a consumer perspective hedonic theory originates from Lancaster's consumer theory (Lancaster, 1966). Each good is considered as a member of group of goods where all the group's goods are used in combination with one another. Thus, the price of the product is not dependent on the amount of characteristics attached to the product but rather on the combination of goods that the buyer decides to buy. From a market perspective, Rosen generalized the approach to include the housing market and integrated it to the general theory. Rosen's model argues that there is a nonlinear relationship between the price of goods and their inherent characteristics. The implicit price is a function of the characteristics associated with the good (Rosen, 1974).

Hedonic Price method's application is mainly based on four key assumptions. First assumption pertains to the homogeneity in the housing market. This assumption is arguable since housing products can differ from one another in terms of locational, structural, neighbourhood or any other attribute and criteria. Second assumption is that the housing market operates under perfect competition. This

is a justifiable assumption since there are many buyers and seller that have freedom to enter and exit the market and consequently no buyer or seller can affect the market. Third assumption pertains to the fact that buyers and sellers have perfect information related to the housing product. Though difficult to achieve, in reality this assumption seems reasonable since buying a house requires a significant amount of capital, the process of buying a house involves collection of a lot of online and offline information before making a purchase. The fourth and final assumption is the assumption of market equilibrium. This assumption is arguable for it is idealistic to assume that the price vector will adjust instantaneously to changes in demand or supply at any point in time thus it does not take into account the relationships between the implicit prices of attributes or characteristics.

HPM has been applied in housing market research (Ball, 1973); (Chau,et al. , 2001); (Freeman, 1979); (Leggett, C. G. & Bockstael, N. E., 2000). HPM was applied in estimation of housing prices in Hong Kong (Chen, 2009). According to this study there is a difference between harbour and mountain view while pricing the houses, namely harbour view, may increase the price while mountain view may decrease it. Keng T.Y (2000) also used HPM to analyze the factors that impact the house price in Malaysia.

The findings are consistent with the hypothesis that house prices are determined by the economic and financial factors identified as relevant for the demand and supply of housing units in this country.

The usual housing characteristics/attributes that are used by HPM are: Structural attributes; location attributes and/or neighbourhood attributes (Herath, 2010). These attributes include both the qualitative and quantitative characteristics.

3.1 STRUCTURAL ATTRIBUTES

Structural attributes have significant effect and relation on the prices of properties. While some studies argue that preferences for structural attitudes may change according to client and may vary over time and between nations (Kohlhase, 1991) (Kohlhase, 1991), other argue that the comparative desirable attributes of a house are valued and reflected in the prices (Ball, 1973) (Ball, 1973). The main attributes that are equally important across nations and are positively related to the price of a house are: the number of rooms and bedrooms (Garrod, 1992)(Garrod, 1992), floor size (Carroll, T. M., Clauretie, T. M. & Jensen, J., 1996) (Rodriguez, M. & Sirmans, C. F., 1994) (Carroll, T. M., Clauretie, T. M. & Jensen, J., 1996)(Rodriguez, M. & Sirmans, C. F., 1994), number of bathrooms, balconies, garage etc., research has also showed that the age of the building is negatively related to the property price due to higher maintenance and repair cost (Clapp, 1998) (Clapp, 1998). On the other hand, attributes like structural quality, heating system, thermal isolation, and basement and so on are difficult and change across nations depending on the traditional values and climate.

4. METHODOLOGY

The location and the surrounding of the property/ house is a very important attribute that has a significant effect on the buying decision. It defines the distance of the house to the city Centre, shopping complexes, job place, school and other related attributes. The main location attributes that have been employed by the studies are the pollution level, the socioeconomic class of the inhabitants, the racial composition, transport modes, the view of the property and the floor level. The main neighbourhood attitudes employed are traffic noise, distance to shopping complexes and the size of shopping complexes, pleasant landscape, pollution, and distance to sport and other recreational facilities.

From the above listed structural and location and neighbourhood attributes due to the lack of available data and difficulty in creations of certain attitudes this study employs only available attitudes for the apartments in Tirana. More precisely, from a structural attributes we use number of rooms, number of bedrooms, number of toilets, size of the apartment in m^2, the existence of balcony and, the existence of elevator. Whereas from a location attributes we use the floor level and location by using 11 mini-districts and the rest as 12th mini-district.
The application of the hedonic method is known for quite some time; it was only in recent years that hedonics has been used in a more extensive way in current statistical production and different analyses. Hedonic pricing methodologies build upon the idea those different characteristics of a good or service impact differently on their evaluation by consumers. Thus, the first stage of this methodology consists of specifying a hedonic price function, i.e. a function relating transaction prices to the relevant characteristics of the good or service.

$$P_{nt} = f(c_{1nt}, c_{2nt}, \ldots, c_{mnt})$$

Where: p is price of goods or services, c1nt, c2nt... cmnt are the characteristics of them.

In hedonic analyses of the housing market, those characteristics typically correspond to individual dwelling and location-related features. Using regression techniques, it is then possible to estimate the parameters in the hedonic price function, which may be interpreted as the implicit marginal prices for each characteristic. Based on the estimated marginal prices, housing prices can be straightforwardly adjusted in order to remove the idiosyncratic influence of those sources of heterogeneity. The particular adjustment carried out depends on the form of the hedonic price function (e.g. linear, log-linear).

The first step to obtain a hedonic index is the specification of a hedonic price function, which is typically linear:

$$P_{it} = x_t + u_t$$

Or log-linear

$$\ln(P_{it}) = x_t \beta_t^* + u_t^*$$

Where x_{it} are the k characteristics (or functions of the characteristics) of house i at period t, βt(β*t) is a (K+1)-vector of parameters (implicit prices) to be estimated, and it (u*it) is the error term, standing for the non-explained part of the price[1] .
This method is based on the assumption that people value the characteristics of a good, or the services it provides, rather than the good itself. Thus, prices will reflect the value of a set of characteristics, including environmental characteristics that people consider important when purchasing the good.

We analyzed only Tirana, capital city of Albania, because the volume of construction of dwellings is higher compare with the other cities, the demand for them is increased year by year. So, Tirana is a representative for the whole country. Based on the characteristics of Albanian housing market we have studied and analysed only the market for apartments. The method used for estimation is the Ordinary Least Squares (OLS) method.

The variables that are included in our model are: age (new or old), number of rooms, number of bedrooms, number of toilets, size in m^2, floor, elevator (dummy variable), balcony (dummy variable), location, mini-districts (administrative separation of Tirana), price per m^2 (the offered price), total price (the offer priced). The numbers of observations are 148; the data are collected for year 2017-2018 from Real Estates agencies in Tirana.

Based on hedonic function our model has this form:

$Lnprice=\beta_0 + \beta_1\ size + \beta_2\ rooms + \beta_3\ bedroom + \beta_4\ toilet + \beta_5\ elevator + \beta_6\ balcony + \beta_7\ minidistricts + \beta_8\ age + \beta_9\ floor$

Where: the sign of the parameters of each variables (features) shows the impact of each of them in price.

5. EMPIRICAL RESULTS

The main empirical results of the employed HPM are shown in Table 2.

From the independent variables measured and tested in the regression analysis the signs of all coefficient results as expected. More specifically, the variables that have expected sign and are statistically significant are: Size in m2, number of bedrooms, age of the building or year of construction, and most of the district variables. The independent variables that indicate positive effect on the price of the house are: size, bedroom, elevator, toilet, age, and district 5 (the most famous and expensive district of Tirana). In addition, the variables that indicate negative effect on the price are: floor level, balcony, and most of the districts except district five. Lastly, the R2 which indicates the coefficient of determination of the model is 74% which shows that the chosen independent variables explain 74.1 percent of the prices of houses in Tirana.

[1] http://ec.europa.eu/eurostat/documents/272892/272983/Detailed_Technical_Manual_on_Owner-Occupied_Housing-v2.pdf/4ccf3133-2309-4316-b7fc-db3b89011b76

Table 1: Population and housing growth, Census 1989, 2001, 2011

Model	β	t	Sig.
(Constant)	11.037145	23.138	0.000
Size	0.008	3.272	0.001
bedroom	0.156	3.210	0.002
floor	-0.039	-0.075	0.931
Elevator	0.009	0.017	0.975
toilet	0.007	0.011	0.555
age	0.103	2.205	0.003
balcony	-0.087	-1.446	0.015
minidistrict3	-0.621	-7.813	0.000
minidistrict13	-0.474	-6.543	0.000
minidistrict6	-0.478	-6.125	0.000
minidistrict11	-0.314	-2.928	0.004
minidistrict4	-0.267	-2.611	0.010
minidistrict5	0.136	2.062	0.041
R^2	**74.1%**		
F-statistics	**47.046**		

6. CONCLUSION

This study empirically examined the predictive power of HPM which determines the intrinsic value of each attribute or characteristic and is also valuable tool for all the different agents that form the real estate market: buyers and sellers, owners and investors, builders and resellers, and banks and fiscal authorities.

It may be the first study that employs HPM for Tirana. The coefficients show that most of the independent variables are highly significant and indicate an expected correct relation with the price. However, some variables though they exerts expected sign they do not seem to be significant.

More specifically, floor level does not seem significant and shows that Albanian household prefers to buy apartments in the lower levels of the building. Similarly, the variable balcony is not significant because it indicates that Albanian household seems to be more rational and ask for more space inside the apartment rather than balconies.

The main limitation of this study which at the same time may also, be a suggestion for future research is the unavailability of data and inability of creation of other location and neighbourhood attributes and variables for Albania in general and capital Tirana in specific.

BIBLIOGRAPHY

BBall, M. (1973). Recent empirical work of the determinants of relative house. Urban Studies, vol. 10.

Carroll, T. M., Clauretie, T. M. & Jensen, J. (1996). Living next to godliness: Residential property values and churches. Journal of Real Estate Finance and Economics, vol. 12,, 319-330.

Chau,et al. . (2001). The pricing of "luckiness" in the apartment market. Journal of Real Estate Literature vo. 9. no 1., 1- 22.

Chen, C. J. (2009). Value of scenic views: Hedonic assessment of private housing in Hong Kong. Landscape and Urban Planning, volume 91, 226-234.

Clapp, J. M. (1998). Residential hedonic models: A rational expectations approach to age effects. Journal of Urban Economics, vol. 44, 415-437.

Freeman, A. M. (1979). Hedonic prices, property values and measuring environmental benefits: A survey of the issues. Scandinavian Journal of Economics vol. 81, 154-171.
Garrod, G. a. (1992). Valuing Goods' Characteristics: An Application of the Hedonic Price Method to Environmental Attributes. Journal of Environmental Management, 34, 59-76.

Herath, S. a. (2010). The hedonic price method in real estate and housing market research. A review of the literature. SRE - Discussion Papers, WU Vienna, University of Economics and Business, Vienna.

INSTAT. (2012). Population and Housing Census 2011. Tirana: INSTAT.

INSTAT. (2014). Albania dwelling and living conditions.Tirana: INSTAT.

Kohlhase, J. E. (1991). The impact of toxic waste sites on housing values. Journal of urban Economics, vol.30, 1-26.

Lancaster, K. J. (1966). A new approach to consumer theory. Journal of Political Economy, 74(1), 132-157.

Leggett, C. G. & Bockstael, N. E. (2000). Evidence of the effects of water quality on residential land prices. Journal of Economics and Management, vol. 39,, 125-141.
Rodriguez, M. & Sirmans, C. F. (1994). Quantifying the value of a view in single-family. Appraisal Journal vol. 62,, 600-603.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, Vol. 82, Jan./Feb., 34-55.

DEVELOPMENT OF SUPPLY, USE AND INPUT-OUTPUT TABLES IN ALBANIA

INPUT-OUTPUT ANALYSIS OF INTER-SECTORIAL LINKAGES

ERMIR LIÇO, INSTITUTE OF STATISTICS
elico@instat.gov.al

Abstract

This article provides a summary on the developments of input-output framework in Albania and an analysis of the economy inter-sectorial linkages, based on the recently published input-output table. Development, compilation and use of input-output tables in analysis and macroeconomic forecast can have important policy implications examining the effects of economic policies.

Using input-output table for the 2013, on this article will investigate multipliers, inter-sectorial linkages and identify key sectors in Albanian economy applying traditional backward and forward linkage methods developed by Chenery-Watanabe. The analyzes are done by dividing input-output table for imported and domestic products for improving the measurement of domestic linkages, as the Albanian economy is small and dependent on imports.

KEY WORDS:

Supply and use tables; Symmetric input-output tables; Multipliers; Backward linkages; Forward linkages; Key sectors

1. INTRODUCTION

The System of National Accounts contains a wide range of macroeconomic indicators. One of the most important indicators is Gross Domestic Product (GDP) which is used for measuring the economic performance and the power of a country. GDP is estimated by three different approaches based on a different view of the economic system and using different data sources. From the theoretical point of view each method should provide the same result, but in reality they may generate different results.

In overall it could not be said which of the methods is most reliable, as it depends on the quality of the data sources used in each of the methods. It is the scope of the balancing process to determine the quality of data sources, by using optimally the existing information to get a more accurate estimate. In other words, balancing process requires that statistical data from the production, expenditure, or income side, should be adjusted and equated between them. This equation is often known as the Supply and Use Tables (SUT), describing the flows of total resources of goods and services produced domestically or imported from suppliers to the users. The final product of balancing process should be a fully balanced set of accounts and articulated with a single GDP estimate.

The construction of the input-output system was inspired by the work carried out by Wassily Leontief in 1930, considered as one of the greatest contributions of economics of the 20th century (Baumol, 2000).

Meanwhile, in 1968, the United Nations published the "System of National Accounts, Studies and Methods", which addressed the integration of these tables into the national accounts system, greatly improving the empirical analysis of the economy. Further developments are also found in updated versions of international standards in the System of National Accounts 1993 and 2008 (SNA 1993 and 2008).

The concept of input-output tables (IOT) was treated as a separate part of the European System of Accounts, 1995 (ESA 1995). It consists of three main tables and the most importance was given to supply and use tables as they provide a clearer picture of product flows. The most recent theoretical treatment, related to these tables, is found in the European System of National Accounts, 2010 (ESA 2010).

One of the advantages of this system, which is also the basis for input-output analysis, is the conceptual simplicity of matrix presentation, analyzing and describing the flow of products and services across industries. Production of an economic activity is interdependent with the production of other activities that use this product as raw materials in production processes or this product uses other branch products as raw materials.

The input-output analysis and the work started by Leontief was further expanded by Rasmussen (1956), Chenery-Watanabe (1958), Hirschman (1958), Gosh, A. (1968), Miller and Blair (1985), Fleissner (1993), Holub and Schnabl (1994), United Nations (1996), Kurz, Dietzenbacher and Lager (1998), Thijs Ten Raa (2006).

Recent developments in the field of input-output analysis relate to empirical analysis that addresses a number of issues for globalization phenomena (Wixted, B., N. Yamano and C. Webb, 2006). Hence, new data requirements and new analytical techniques have been increased in the global concept and to assess the effects on the economy of phenomena such as global production, multinational enterprises, factoryless manufacturing, etc.

In a world with more competitive global markets and a more interdependent production process, such as the production of a smartphone, whose processor is manufactured in the USA, the screen in South Korea and the assembly process is carried out in China, has created a specialization in production, diversification and greater complexity of products.

The interdependence between sectors and countries can be described with a linear equation that expresses the balance between total inputs and outputs for each good and service generated for a country's entire economy or expansion of input-output analysis in the global concept. The production structure for each of the sectors of the economy is presented by technical coefficients that describe in quantitative terms the relationship between inputs used and produced output.

The following sectorial linkage analysis are based on the input-output table for domestic production, with an emphasis on interdependence between countries, thus whether the inputs used are from domestic or imported output. Taking into account this division, a more accurate estimate of economic effects is given, as the impact of domestically produced inputs is greater and, as the Albanian economy is a small and highly dependent economy, it may overestimate the multiplier effect for a given sector.

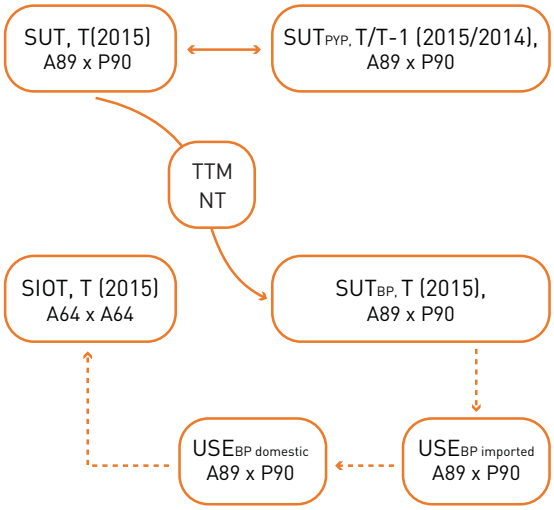
2. INPUT-OUTPUT FRAMEWORK STRUCTURE

Input-output framework is an integral part of the European System of Accounts (ESA 2010) and fully consistent with the world-wide System of National Accounts (SNA 2008).
The input-output framework consists of three types of tables:
- supply tables,
- use tables and
- symmetric input-output tables.

The compilation of supply-use and input-output tables does not have a long tradition in Albania and INSTAT has included in its activity the development of this system under IPA 2007 (IPA 2007 Multi-beneficiary Statistical Cooperation Program, National Accounts). In February 2015, for the first time were published supply and use tables at current prices for 2009-2011 and derived symmetric input-output table (SIOT) for 2011. During years 2016 and 2017, supply and use tables and input-output for 2012-2014 were published. The compilation and publication of the tables is done three years after the reference year (t + 3), in line with Eurostat's methodology and requirements.
The input-output system base is compiling and balancing supply and use tables at current prices. The compilation of the components is done at a detailed level and then aggregated at the level of 89 activities, according to the Nomenclature of Economic Activities Rev. 2 (NACE Rev. 2) and 90 products according to the Classification of Product by Activity (CPA 2008). This is the level at which balancing is performed as well.
At this point begins the compilation of supply and use tables at constant prices (SUTPYP) where the composite components at a detailed level of two consecutive years are deflated with specific indices at product level and performed a simultaneous balancing process.
Another important table of the system is the supply and use table at basic prices (SUTBP), which is compiled from a balanced supply and use table at current prices, by subtracting the trade and transport margins (TTMs) and taxes net on products (NT) from use side. The use table at basic prices is breakdown for domestic production and imports is the basis for the conversion of symmetric input-output tables (TSIOs).

[1] Previous year prices

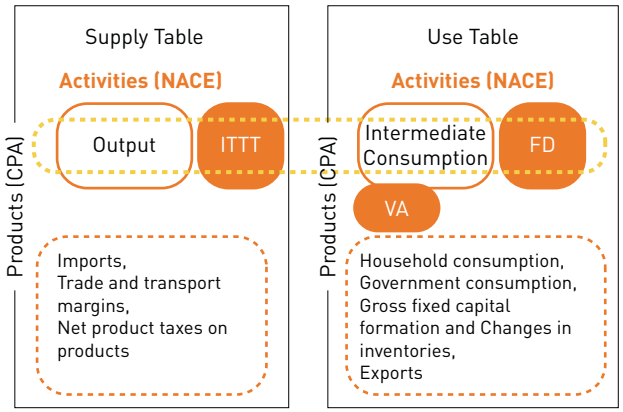
Figure 1: Input-Output framework structure



2.1 SUPPLY AND USE TABLES

The system of SUT is set up by two main tables. On the left side are total resources of a country, produced or imported at product level, while on the right side are presented the uses for each product. Table balancing expresses the identity of the reconciliation of economic resources, produced or imported, with flows by uses side as intermediate consumption in production processes, final consumption or exports.

Figure 2: Supply and use table's structure



2.1.1 Supply Table

The supply table contains three main matrices:
- Production matrix,
- Import matrix,
- Adjustment matrix (taxes and trade and transport margins).

The production matrix shows the production of industries by products at basic prices, presenting the primary production in diagonal, according to the definition that an enterprise is classified in an industry by the product it produces and off diagonals secondary products.

The matrices have a structure in rows by category of products, allowing the horizontal aggregation of all elements by the total output of industries at basic prices, to the total supply at purchaser prices. Firstly, domestic product at basic prices and imports in CIF value are aggregated to the total supply at basic prices. In the second step of the evaluation vectors of trade and transport margins and net taxes on products are added to the total supply at basic prices in order to obtain the total supply at purchaser prices.

2.1.2 Use Table

A use table shows, by industries (in columns), the intermediate consumption structure used in the production process, along with the primary inputs in the last quadrant (compensation of employees, net operating surplus, other taxes less subsidies on production, and consumption of fixed capital). By product (in rows), the use table shows the demand for intermediate consumption in production and final consumption (household final consumption, government consumption, NIPSH's consumption and gross capital formation) or exports.
The use table is valued at purchasers' prices, which also represents the actual paid value, unlike the supply that is valued at basic prices.
Depending on whether it will be compiled a table of uses at basic prices or at purchasers' prices, it either subtracted or added to the flow of trade and transport margins and net taxes matrices.

2.1.3 Balancing supply and use tables

Balancing process plays a crucial role on the Supply and Use table compilation. After a detailed elaboration of each component, all resources must be faced with intermediate and final uses data. Before analyzing discrepancies at product level, the statistical discrepancy between the different methods of GDP estimation is analyzed. In the system of supply and use tables, this statistical discrepancy should be eliminated, and therefore macroeconomic balancing is required.
In the first stage of balancing process, where discrepancies are between 5% and 10%, they will be manually analyzed and balanced. In cases where discrepancies are greater than 10% it is required a treatment in the basic source of information. In both cases data sources are analyzed to determine the exact cause of the differences. During this phase it may be necessary that different components of supply or uses be reclassified or re-estimated.

The next stage is an automatic balancing process when the plausible economic pictures have been defined for all the economic components, an algorithm is used for final balancing of products with a discrepancy between supply and use of less than 5 %. For this final balancing process is used an algorithm based on the RAS[2] method of reaching a balanced and consistent supply and use table for all industries (columns) and products (rows).

2.2 TRANSFORMATION TO INPUT-OUTPUT TABLES

The conversion of SUT into IOT consists on removing secondary output from the output matrix and respective inputs from the intermediate consumption matrix. There are 4 possibilities depending on whether the secondary output moves to a different industry (left/right) or to a different product (up/down) and whether the input structure is used from the point of origin or from the point of destination.
The data base for the conversion of symmetric input-output tables from supply and use tables comprises the following tables:
• Supply tables at basic prices
• Use table at basic prices
• Use tables for domestic output at basic prices
• Use tables for imports at basic prices
In a supply system at basic prices, the columns for trade and transport margins and net taxes on products become irrelevant in the supply table as the valuation matrices were deducted from the use table at purchasers' prices. However, non-deductible taxes less subsidies on products form an additional row in the use tables, as total uses continue to be valued at purchasers' prices.
The following section will explain in detail four basic models of compiling symmetric input-output tables. These include two models that are based on the assumption of a technology which generates input-output tables, product-by-product. In this case, the input-output tables are composed of homogeneous products by rows and homogeneous units of production (industries) by columns.
Two other models are based on the assumption of a fixed sales structure from where industry-to-industry IOT are derived. The results of these models are input-output tables with industries by rows and industries by columns.
The four basic transformation models used for compiling product-to-product or industry-to-industry IOT are based on the following assumptions:
a) Product x Product
• Product technology assumption (Model A).
Each product is produced in its own specific way, irrespective of the industry where it is produced. So, technology used to produce a specific product is the same for all industries.
• Industry technology assumption (Mode B). Each

2 The acronym indicates that the coefficient update is done by pre-multiplying and post-multiplying the matrix of A-matrix technical matrices with R and S matrices (Stone and Brown 1965)

industry has its own specific way of production, irrespective of its product mix. In this model, the economic activity does not change regardless the structure of the product that is producing.
b) Industry x Industry
• Fixed industry sales structure assumption (Model C). Each industry has its own specific sales structure, irrespective of its product mix.
• Fixed product sales structure assumption (Model D). Each product has its own specific sales structure, irrespective of the industry where it is produced.
In addition to the basic models, there are two other transformation models:
• The hybrid technology assumption. The hybrid technology assumption combines the product technology assumption and the industry technology assumption to avoid negatives in product-by-product input-output tables.
• The Almon procedure. The Almon procedure is a mathematical algorithm designed for compiling product-by-product input-output tables which are based in essence on the product technology assumption but avoids step-by-step procedure negatives in the derived input-output tables.
In a summarized way the mathematical derivatives are presented in the Eurostat Manual for the Supply-Use and Input-Output Tables (Eurostat 2008), for the symmetrical input-output table derivation (SIOT). Regardless of the selected method, the input-output symmetric table will not be fully consistent with the balanced supply and use table. As a result, the above model is modified and the differences are distributed to create a new use table. INSTAT published IOT industry-by-industry type based on the assumption of the fixed product sales structure (Model D), a model that does not give negative values. Besides, industry-by-industry tables provide more analysis opportunities as well.

Table 1: Summary of transformation models

	MODEL A Product technology Product-by-product	MODEL B Industry technology Product-by-product	MODEL C Fixed industry sales structure Industry-by-industry	MODEL D Fixed product sales structure Industry-by-industry
Transformation matrix	T = inv (V')*diag(q)	T = inv[diag(g)] * V	T = diag(g) * inv (V')	T = V * inv[diag(q)]
Input coefficients	A = U*T* inv[(diag(q)]	A = U*T* inv[(diag(q)]	A = T * U * inv[(diag(g)]	A = T * U * inv[(diag(g)]
Intermediates	S = U * T	S = U * T	B = T * U	B = T * U
Value added	E = W * T	E = W * T	W = W	W = W
Final demand	Y = Y	Y = Y	F = T * Y	F = T * Y
Output	q = inv(I - A) * y	q = inv(I- A) * y	g = inv(I - A) * y	g = inv(I - A) * y

Source: Eurostat Manual of Supply, Use and Input-Output Tables, page 351

3. INTER-SECTORAL LINKAGES ON ALBANIAN ECONOMY

3.1 LITERATURE REVIEW

Input-Output analyses are linked with static input-output system of Wassily Leontief, based on the linear Leontief production function. Linkage analysis, used to examine the interdependency of production structures, was introduced by the works of Rasmussen (1956) and then important contributions have given Chenery & Watanabe (1958) and Hirschman (1958).

Balance between total inputs and output is described by the linear function:

$$\begin{aligned} a_{11} X_1 + a_{12} X_2 + ... + a_{1n} X_n + Y_1 &= X_1 \\ a_{21} X_1 + a_{22} X_2 + ... + a_{2n} X_n + Y_2 &= X_2 \\ a_{n1} X_1 + a_{n2} X_2 + ... + a_{nn} X_n + Y_n &= X_n \end{aligned} \quad (1)$$

We assume that all sectors produce with fixed proportions in relation to output and changing factor prices have no influence on the technical input coefficients. The computation of the input technical coefficients matrix is described by the following mathematical form:

$$a_{ij} = \frac{x_{ij}}{x_j} \quad (2)$$

Where:
a_{ij} = linkage coefficients,
x_{ij} = output from sector i used to sector j,
x_j = output of sectors j.

Another method for the calculation of input coefficients is the transformation of equation 2 to matrix form:

$$AX + Y = X \quad (3)$$

Matrix A defines the technical input–output coefficients, vector X present the output while vector Y represent the exogenous aggregate of final uses.

$$\begin{aligned} X - AX &= X \\ (I - A)\, X &= Y \\ X &= (I - A)^{-1}\, Y \end{aligned} \quad (4)$$

Where I is the identity matrix and inverted $(I - A)^{-1}$ is the so-called Leontief inverse.

$$(I - A)^{-1} = \begin{bmatrix} 1 - a_{11} & -a_{12} \\ a_{21} & 1 - a_{22} \end{bmatrix} \quad (5)$$

The inputs structure represented by matrix A and inverse (I – A)-1 are the foundations on which are constructed the input-output models. These matrixes denote the existing linkages between the industries in economy and the effect on output from the changes in final demand.
Another family of I-O models also exists which is based on output coefficients, developed by Ghosh (Ghosh 1958). The use-side Leontief models and the supply-side Ghosh models can be used to study the impact of changes in final use and primary inputs on output as well as price and cost effects. The dual character of Leontief models and Ghosh models is discussed in Oosterhaven (1996), Dietzenbacher (1997), De Mesnard (2009) and Rueda-Cantuche (2011). Leontief and Ghosh models are similar but opposite in structure, almost as mirror images of each other. Leontief models use fixed input coefficients whilst the Ghosh models rely on fixed output coefficients.

3.2 METHODOLOGY

Inter-industrial linkage analysis has important policy implications and is used to determine the weight of each producing of goods and services industry. This analysis is one of central method for macroeconomic modelling. In the following section we will present in a summarized way the concepts that our analysis is based.

3.2.1 Multipliers

Multipliers estimate how the effects of the change of conditions for an industry are transmitted to the total economy. Essentially, multipliers estimate the total 'gross' effect generated in economy as change by one unit of the exogenous final demand of a given sector on:

- outputs of the sectors in the economy;
- value added and income earned by the households;
- employment that is expected to be generated by the new activity levels.

Output multiplier, for an industry j is defined as the total value of production in all industries of the economy that is necessary for all stages of production in order to produce one unit of product j for final use. The output multipliers (Oj) are equal to the column sum of the Leontief Inverse matrix and are defined as:

$$O_j = \sum_1^n a_{ij} \quad (6)$$

Income multipliers, as the output multipliers, attempt to identify the impacts of final use changes on income received by households for labour supply. The direct and indirect requirements for wages which are incorporated in one unit of output for final use is estimated by the equation:

$$Z = B(I - A)^{-1} \quad (7)$$

B = vector of input coefficients for wages,
I = unit matrix,
A = matrix of input coefficients for intermediate consumption,
Z = vector with results for direct and indirect requirements for wages.

Employment multipliers are calculated using the physical labour input coefficients. They show the direct and indirect demand for labour as result to final demand change. The Employment multipliers are calculated using the following equation:

$$Z = E(I - A)^{-1} \quad (8)$$

E = matrix of input coefficients for labour (physical persons number per million ALL of output),
Z = matrix with results for direct and indirect requirements for labour (physical persons number).

3.2.2 Backward and forward linkages

In input-output analysis and framework, the production growth by a particular industry has two kinds of effects on other industries in the economy. If an industry j increases its output, more inputs are required including more intermediate consumption from other industries. Also, the industries that use this product as input have the possibility to increase their production.

The term 'backward linkage' is used to indicate the inter-connection of a particular industry to other industries from which it purchases inputs (use side). On the other hand, increased output of industry j request b*j units of output for the total economy, consisting in one unit for j and additional amounts of direct and indirect inputs used by other industries. This is equal to output multiplier and estimate the effect of change in final demand by one unit in total output of industries. The term 'forward linkage' is used to indicate the interconnection of a particular industry to those to which it sells its output. Forward linkages analyze the relation between output and primary inputs, analyzing the supply side.
The Leontief quantity model will help to identify backward linkages while the Ghosh price model can be used to identify forward linkages.
Forward and backward linkages analysis is based on model Chenery and Watanabe (1958). The CW (Chenery-Watanabe) backward linkages are equal to the sums of the inverse input coefficient matrix and are defined as follows:

$$BL^{CW}{}_j = \sum_1^n a_{ij} \qquad (9)$$

Where:

$BL^{CW}{}_j$ denotes the backward linkage of sector j,
a_{ij} denotes the input coefficient matrix.

The forward linkages CW are equal to the sums of rows of the output coefficient, matrix B. The CW forward linkages of sector i are defined as:

$$FL^{CW}i = \sum_1^n b_{ij} \qquad (10)$$

Where:

$FL^{CW}i$ denotes the forward linkage of sector i,
b_{ij} denotes the output coefficient of sector i to sector j.

The Chenery-Watanabe method is based on direct input or output coefficients, measuring only the effects generated by the inter-relationships between sectors. If the linkages are greater than 1, the sector has a higher impact on economy. Also are estimated the normalized linkages divided by the total linkages average.

3.2.3 Key Sectors in Economy

Key sectors analysis is based on normalized backward and forward linkages. The normalized linkage indicators are calculated by using following formulas:

$$NBLj = \sum_1^n a_{ij} / \frac{1}{n}\sum_1^n a_j$$
$$NFLi = \sum_1^n b_{ij} / \frac{1}{n}\sum_1^n b_j \qquad (11)$$

If NBL>1, then the growth by one unit of final demand for a given j sector will generate an impact greater than 1 in economy. In the same way, if NFL>1, then the growth by one unit of primary inputs availability for a given sector i will generate a growth greater than 1 on total economy. We classify a sector, as key sector if NBL>1 and NFL>1, with strong backward linkages if NBL>1 and NFL<1, and with strong forward orientation if NFL>1 and NBL<1.

For the calculation of weighted linkages is used, for the backward linkages the specific weigh of final demand, identifying that this linkages are demand driven. Regarding the forward linkages, is weighing using the share of a given sector's value added to total value added in the economy reflecting supply-driven linkages model.

3.3 RESULTS

The inter-sectorial linkages analysis is based on the results of industry by industry input-output tables for year 2013 transformed from supply and use tables based on Model D (the assumption of a fixed sales structure for the product). The analysis is done at the level of detail 35*35 according to the Economic Activity Nomenclature (NACE Rev. 2) presented in Annex 1. Multipliers estimate that a change of conditions in a given sector are transmitted to the economy and are the basis for the impact analysis. Multipliers for output, value added compensation of employees and employment are presented in Annex 2. Multipliers with values greater than average, in conditions of change in final demand for products of these industries, will have a greater impact on the total economy. To analyze the economic impact deriving from an increase of 5%, in the final demand for Construction (F – Residential and non-residential constructions) is presented in Annex 4 the direct effect of change in output, imports, value added and household income. The highest impact is in the Construction Sector which has been directly impacted, but an indirect impact generated also in the linked sectors and the sum of direct and indirect give the total impact on economy.
The backward and forward linkages analysis is based on the Chenery-Watanabe method and

results are presented in Annex 3. The analysis of key sectors of the economy is based on the normalized backward and forward linkages; firstly are estimated the unweighted linkages and then are estimated the weighted linkages. The letters K, G and B indicates key sectors, strong forward linkages and strong backward linkages.

Based on the ranking of absolute value for backward and forward linkages, which are presented in Annex 3, the sectors with strong unweighted forward linkages are: Manufacture of coke and refined petroleum products (C19), Manufacture of rubber and plastic products and other non-metallic mineral products (C22_C23), Sewerage, waste management and remediation activities (E37_E39), Manufacture of wood and paper products, and printing (C16_C18), Manufacture of basic metals and fabricated metal products, except machinery and equipment (C24_ C25).

Industries with strong unweighted backward linkages are: Manufacture of coke and refined petroleum products (C19), Construction (F), Telecommunication (J61), Land transport and transport via pipelines (H49).
On the other hand, if we take into account weighted linkages, the industries with strong backward linkages are estimated: Agriculture, Forestry and Fishing (A01_03), Construction (F), Wholesale trade, except of motor vehicles and motorcycles (G46), Mining and quarrying (B), Retail trade, except of motor vehicles and motorcycles (G47).
Industries that present strong forward orientation based on weighted linkages are: Construction (F), Agriculture, Forestry and Fishing (A01_03), Telecommunication (J61), Wholesale trade, except of motor vehicles and motorcycles (G46), Accommodation and food service activities (I).

3.4 LIMITATIONS OF EMPIRICAL ANALYSIS BASED INPUT-OUTPUT ANALYSIS

As noted in the previous sections, I-O models are consistency models, capable of assessing the main elements of interdependence between the economic sectors and the impact of changes in economic conditions and external shocks in the economy. Although analytical techniques based on I-O models have been used for a long time, they also present limitations that are documented by studies conducted in this area (ABS, 2012).

Firstly, these models tend to be extremely demand driven and do not take into account supply-side constraints. It is assumed that an increase in demand for an additional unit of output in one industry will lead to increased production regardless of sources for another sector. In other words, it is assumed that the industry resource curve is perfectly elastic.

Secondly, prices are exogenously defines in these models and it is assumed that prices are not affected by economic policies or external shocks.

Thirdly, the assumption of fixed coefficients or the assumption of the fixed input structure used in I-O models is not realistic, as it excludes the technological effects of different industries.

Fourthly, the lack of any relationship between the constraints of the primary factors and the final demand is another shortcoming of these models. Industrial output functions are simple linear functions related with the final demand.

Fifthly, these models do not allow consumers to change their budget shares on consumption goods when their income changes.

In conclusion, dealing with international trade is incomplete in these models. As previously noted, exports, together with the final demand, are exogenous and imports are treated as uncompetitive.

4. CONCLUSIONS

This article provides a summary of the input-output system developments in Albania and an analysis of the inter-sectorial linkages of the economy, based on these tables. In order to analyze the inter-sectorial linkages are used the input-output tables for year 2013, divided for domestic products and imports, to improve the measurement of linkages, analyzing only the domestic flow, since the Albanian economy is small and dependent on imports. Applying the traditional method of backward and forward linkages developed by Chenery-Watanabe for weighted linkages, key sectors in the Albanian economy are estimated in the following industries: Agriculture, Forestry and Fishing (A01_03), Construction (F), Wholesale trade, except of motor vehicles and motorcycles (G46), Mining and quarrying (B), Retail trade, except of motor vehicles and motorcycles (47), Education (P85), Public administration and defence; compulsory social security security(O84). The impact of these sectors, both by the demand side by increasing the demand for the products of these sectors, also by the supply of industries that use the products of these sectors as inputs in production processes, will generate the greater impact on economy.
The question that arises after analyzing the results of the input-output analysis is that which of the

estimates, based on un-weighted or weighted linkages, gives a more accurate result. To address this issue, we have analyzed the two sectors that have the highest total linkages according to both estimates; Sector for Agriculture, Forestry and Fishing (A01_03) and Manufacture of coke and refined petroleum products (C19). Based on the un-weighted links, the Sector for Agriculture, Forestry and Fishing (A01_03) is estimated with weak forward and backward linkages as a result of underdevelopment of related sectors. If we analyze the forward linkages, an increase of output of this sector would not lead to an increase of the production from the sectors that use these products as inputs, as are not developed the associative processing industries. In the same line, if we analyze the forward linkages, an increase of demand for products of the Agriculture, Forestry and Fishing Sector would not affect the industries from which this sector buys the inputs, given that most of these products are imported. On the other hand, if we analyze the weighted links for this sector, as a result of the weight of value added and the final demand, it is considered as the sector with the strongest backward and forward linkages.

If we analyze the Manufacture of coke and refined petroleum products (C19), identified as the sector that is ranked with strong unweighted linkages, consequently, should generate a high impact on the economy, but if we analyze the weighted linkages, this sector is identified with weak linkages as result of the weight of value added and the final demand to the total economy.

To sum up, the result of both methods, even though different, gives valuable information, based on different perspectives. If we analyze the impact with a unit in the Manufacture of coke and refined petroleum products (C19), we expect a high impact as a result of strong absolute backward and forward linkages, but the specific weight of this sector in total value added and the final demand of the economy, is small and as a result the total impact as well. Development of this sector and fulfilment of the demand from domestic production would increase the specific weight of this sector and, as a result, the overall impact on economy.

On the other hand, the Agriculture, Forestry and Fishing Sector has a high specific weight in total of value added and final demand, and have the higher total effect, if we would have an increase in the final demand for the products of this sector. This impact would be higher if the linked sectors were developed, so the coefficients of backward and forward linkages would be greater.

5. ANNEX

Annex 1: Input-Output industry classification by NACE Rev. 2

	Code	Description
1	A01_03	Agriculture, forestry and fishing
2	B	Mining and quarrying
3	C10_C12	Manufacture of food products, beverages and tobacco products
4	C13_C15	Manufacture of textiles, wearing apparel and leather products
5	C16_C18	Manufacture of wood and paper products, and printing
6	C19	Manufacture of coke and refined petroleum products
7	C20_C21	Manufacture of chemical and pharmaceutical products
8	C22_C23	Manufacture of rubber and plastic products and other non-metallic mineral products
9	C24_C25	Manufacture of basic metals and fabricated metal products, except machinery and
10	C26_C30	Manufacture of machinery and equipment
11	C31_C33	Manufacture of furniture; other manufacturing; repair and installation of machinery and
12	D35	Electricity, gas, steam and air-conditioning supply
13	E36	Water supply
14	E37_E39	Sewerage, waste management and remediation activities
15	F	Construction
16	G45	Wholesale and retail trade and repair of motor vehicles and motorcycles
17	G46	Wholesale trade, except of motor vehicles and motorcycles
18	G47	Retail trade, except of motor vehicles and motorcycles
19	H49	Land transport and transport via pipelines

Code		Description
20	H50_H52	Water and air transport; warehousing
21	H53	Postal and courier activities
22	I	Accommodation and food service activities
23	J58_J60	Publishing, audiovisual and broadcasting activities
24	J61	Telecommunications
25	J62_J63	Computer programming, consultancy and related activities; information service activities
26	K64_K66	Financial and insurance activities
27	L68	Real estate activities
28	M69_M71	Legal and accounting activities; management consultancy activities; architectural and
29	M72_M75	Scientific research and development; other professional, scientific and technical activities
30	N77_N82	Administrative and support service activities
31	O84	Public administration and defence; compulsory social security
32	P85	Education
33	Q86_Q88	Human health activities
34	R90_R93	Arts, entertainment and recreation
35	S94_98	Other services and activities of households

Annex 2: Multipliers for Output, Value added, Compensation of employees, Employment

Economic Activity		Output Multiplier			VA Multiplier			CE Multiplier			Employment Multiplier		
		Multiplier	Index	Rank	Multiplier	Index	Rank	Multiplier	Index	Rank	Multiplier	Index	Rank
1	A01_03	1.35	0.94	23	0.99	1.13	1	0.05	0.15	35	1.61	2.43	1
2	B	1.26	0.87	28	0.88	1.00	19	0.21	0.60	34	0.26	0.39	34
3	C10_C12	1.27	0.89	27	0.84	0.96	25	0.21	0.60	33	1.07	1.62	6
4	C13_C15	1.46	1.02	13	0.88	1.00	18	0.44	1.27	8	0.98	1.49	7
5	C16_C18	1.34	0.94	24	0.85	0.97	24	0.32	0.92	18	0.74	1.12	13
6	C19	2.06	1.43	1	0.79	0.90	32	0.29	0.84	24	0.56	0.85	17
7	C20_C21	1.11	0.77	34	0.81	0.93	29	0.32	0.93	17	0.47	0.71	20
8	C22_C23	1.51	1.05	10	0.77	0.89	34	0.25	0.72	31	0.43	0.65	24
9	C24_C25	1.39	0.97	20	0.78	0.89	33	0.22	0.64	32	0.43	0.65	23
10	C26_C30	1.39	0.97	21	0.81	0.93	31	0.30	0.87	22	0.61	0.92	15
11	C31_C33	1.33	0.93	25	0.82	0.93	28	0.33	0.95	14	1.23	1.87	4
12	D35	1.25	0.87	30	0.93	1.07	7	0.27	0.77	28	0.26	0.39	33
13	E36	1.43	0.99	18	0.88	1.01	16	0.47	1.35	6	0.91	1.39	9
14	E37_E39	1.61	1.12	8	0.72	0.83	35	0.27	0.79	26	0.41	0.61	25
15	F	1.81	1.26	2	0.84	0.96	26	0.26	0.74	30	0.45	0.68	22
16	G45	1.10	0.77	35	0.87	0.99	21	0.30	0.86	23	1.46	2.21	2
17	G46	1.38	0.96	22	0.92	1.06	8	0.26	0.74	29	0.30	0.46	32
18	G47	1.28	0.89	26	0.92	1.05	9	0.31	0.88	21	1.22	1.85	5
19	H49	1.69	1.18	4	0.87	1.00	20	0.33	0.94	15	0.58	0.87	16

Economic Activity		Output Multiplier			VA Multiplier			CE Multiplier			Employment Multiplier		
		Multiplier	Index	Rank	Multiplier	Index	Rank	Multiplier	Index	Rank	Multiplier	Index	Rank
20	H50_H52	1.43	1.00	16	0.81	0.93	30	0.27	0.78	27	0.39	0.59	26
21	H53	1.63	1.13	7	0.88	1.00	17	0.48	1.39	5	0.45	0.69	21
22	I	1.48	1.03	12	0.91	1.04	11	0.32	0.93	16	1.35	2.05	3
23	J58_J60	1.64	1.14	6	0.91	1.04	12	0.43	1.24	10	0.55	0.83	18
24	J61	1.74	1.21	3	0.85	0.97	23	0.31	0.89	20	0.37	0.56	28
25	J62_J63	1.54	1.07	9	0.91	1.05	10	0.42	1.20	11	0.32	0.48	31
26	K64_K66	1.46	1.01	14	0.95	1.09	5	0.57	1.64	4	0.37	0.56	29
27	L68	1.21	0.84	32	0.98	1.12	2	0.29	0.84	25	0.06	0.10	35
28	M69_M71	1.42	0.99	19	0.85	0.98	22	0.36	1.02	12	0.33	0.50	30
29	M72_M75	1.49	1.04	11	0.83	0.95	27	0.33	0.95	13	0.47	0.72	19
30	N77_N82	1.43	1.00	17	0.89	1.01	15	0.43	1.24	9	0.38	0.57	27
31	O84	1.23	0.86	31	0.94	1.07	6	0.61	1.75	1	0.95	1.43	8
32	P85	1.17	0.81	33	0.96	1.10	3	0.59	1.71	2	0.91	1.37	10
33	Q86_Q88	1.25	0.87	29	0.89	1.02	14	0.59	1.70	3	0.72	1.08	14
34	R90_R93	1.45	1.01	15	0.95	1.09	4	0.31	0.91	19	0.75	1.13	12
35	S94_98	1.65	1.15	5	0.91	1.04	13	0.44	1.27	7	0.77	1.17	11
Average		**1.44**			**0.8736**			**0.3475**			**0.6602**		

Source: Author's calculation based on INSTAT data.

Annex 3: Backward linkages, Forward linkages and Key Sectors on Economy

Economic Activity		Direct forward	Total forward	Rank	Direct backward	Total backward	Rank	Unweighted			Weighted		
								Total forward	Total backward	Effect	Total forward	Total backward	Effect
1	A01_03	0.25	1.33	26	0.25	1.35	23	0.82	0.94		7.25	5.55	K
2	B	0.39	1.68	15	0.19	1.26	28	1.04	0.87	F	2.31	1.21	K
3	C10_C12	0.12	1.16	29	0.20	1.27	27	0.72	0.89		0.27	0.93	
4	C13_C15	0.12	1.16	28	0.32	1.46	13	0.72	1.02	B	0.47	0.93	
5	C16_C18	0.76	2.27	4	0.24	1.34	24	1.41	0.94	F	0.25	0.11	
6	C19	0.87	2.38	1	0.82	2.06	1	1.48	1.43	K	(0.01)	0.27	
7	C20_C21	0.53	1.84	11	0.08	1.11	34	1.14	0.77	F	0.07	0.35	
8	C22_C23	0.87	2.38	2	0.32	1.51	10	1.48	1.05	K	0.58	0.15	
9	C24_C25	0.65	2.12	5	0.27	1.39	20	1.32	0.97	F	0.54	0.54	

	Economic Activity	Direct forward	Total forward	Rank	Direct backward	Total backward	Rank	Unweighted			Weighted		
								Total forward	Total backward	Effect	Total forward	Total backward	Effect
10	C26_C30	0.27	1.43	21	0.26	1.39	21	0.89	0.97		0.05	1.23	B
11	C31_C33	0.27	1.38	23	0.24	1.33	25	0.86	0.93		0.14	0.26	
12	D35	0.36	1.53	18	0.17	1.25	30	0.95	0.87		0.88	0.49	
13	E36	0.53	1.66	16	0.29	1.43	18	1.03	0.99	F	0.17	0.08	
14	E37_E39	0.75	2.29	3	0.42	1.61	8	1.42	1.12	K	0.23	0.15	
15	F	0.32	1.49	20	0.52	1.81	2	0.92	1.26	B	4.27	7.22	K
16	G45	0.54	1.82	12	0.07	1.10	35	1.13	0.77	F	0.29	0.08	
17	G46	0.34	1.50	19	0.26	1.38	22	0.93	0.96		2.65	1.62	K
18	G47	0.25	1.36	24	0.19	1.28	26	0.85	0.89		1.55	1.06	K
19	H49	0.40	1.61	17	0.48	1.69	4	1.00	1.18	B	0.69	0.99	
20	H50_H52	0.26	1.41	22	0.29	1.43	16	0.87	1.00	B	0.51	1.06	B
21	H53	0.64	1.95	9	0.43	1.63	7	1.21	1.13	K	0.18	0.08	
22	I	0.10	1.14	30	0.35	1.48	12	0.71	1.03	B	0.55	1.40	B
23	J58_J60	0.48	1.72	14	0.42	1.64	6	1.07	1.14	B	0.35	0.27	
24	J61	0.18	1.26	27	0.49	1.74	3	0.79	1.21	B	0.65	1.72	B
25	J62_J63	0.63	1.95	10	0.34	1.54	9	1.21	1.07	K	0.12	0.06	
26	K64_K66	0.70	2.06	7	0.31	1.46	14	1.28	1.01	K	1.29	0.30	F
27	L68[1]	0.23	1.36	25	0.13	1.21	32	0.85	0.84		2.42	1.19	
28	M69_M71	0.68	2.07	6	0.29	1.42	19	1.29	0.99	F	1.03	0.32	F
29	M72_M75	0.54	1.82	13	0.32	1.49	11	1.13	1.04	K	0.20	0.14	
30	N77_N82	0.62	1.96	8	0.29	1.43	17	1.22	1.00	F	1.18	0.45	F
31	O84	0.08	1.10	31	0.16	1.23	31	0.68	0.86		1.23	1.15	K
32	P85	0.04	1.06	32	0.11	1.17	33	0.66	0.81		1.33	1.09	K
33	Q86_Q88	0.02	1.03	34	0.17	1.25	29	0.64	0.87		0.74	0.97	
34	R90_R93	0.02	1.03	35	0.31	1.45	15	0.64	1.01	B	0.24	0.73	
35	S94_98	0.02	1.03	33	0.45	1.65	5	0.64	1.15	B	0.31	0.85	
	Average	**0.40**	**1.61**		**0.30**	**1.44**							

Note: K = Key Sectors, F = Forward linkages, B = Backward linkages.

Source: Author's calculation based on INSTAT data.

Annex 4: Impact analysis due to change in domestic demand for Construction by 5%

Economic Activity	*Domestic Demand Change*			Output change	Change in Imports	Change in Value Added (GDP)	Change in Household Income	Employment change (Number)	*Cumulative Output*				
	%	Value	Total Change						First Round	Second Round	Third Round	Fourth Round	Fifth Round
A01_03	0.00%	-	-	117	34	84	2	145	39	81	103	112	115
B	0.00%	-	-	1,095	51	592	108	84	657	934	1,040	1,077	1,089
C10_C12	0.00%	-	-	24	15	6	2	8	12	19	22	23	24
C13_C15	0.00%	-	-	10	28	5	3	7	1	6	8	9	9
C16_C18	0.00%	-	-	137	169	51	22	49	80	116	130	135	136
C19	0.00%	-	-	34	741		2	9	0	20	29	32	34
C20_C21	0.00%	-	-	26	337	9	4	5	14	22	25	25	26
C22_C23	0.00%	-	-	1,497	296	389	116	196	1,055	1,368	1,457	1,484	1,493
C24_C25	0.00%	-	-	584	516	138	29	80	365	506	558	575	581
C26_C30	0.00%	-	-	34	371	15	6	13	21	29	32	33	34
C31_C33	0.00%	-	-	18	11	6	3	16	4	12	16	17	18
D35	0.00%	-	-	156	38	113	28	22	75	126	145	152	155
E36	0.00%	-	-	19	1	10	6	14	11	16	18	18	19
E37_E39	0.00%	-	-	220	93	48	24	23	118	183	207	216	218
F	**5.00%**	**-**	**14,970**	**18,485**	**986**	**5,818**	**1,572**	**3,053**	**17,296**	**18,123**	**18,370**	**18,448**	**18,473**
G45	0.00%	-	-	59	2	38	13	77	25	45	54	57	58
G46	0.00%	-	-	594	12	367	89	54	354	509	565	584	590
G47	0.00%	-	-	308	15	218	71	333	201	273	297	304	307
H49	0.00%	-	-	303	85	105	34	95	165	253	286	297	301
H50_H52	0.00%	-	-	88	71	37	12	13	24	60	78	84	87
H53	0.00%	-	-	63	4	28	20	11	33	51	58	61	62
I	0.00%	-	-	34	15	16	8	31	16	26	31	33	33
J58_J60	0.00%	-	-	15	9	7	3	4	2	7	11	13	14
J61	0.00%	-	-	89	22	29	9	6	40	67	81	86	88
J62_J63	0.00%	-	-	18	1	8	4	1	2	10	15	17	17
K64_K66	0.00%	-	-	336	29	207	143	77	160	266	311	327	333
L68	0.00%	-	-	249	0	212	59	-	121	198	230	243	247
M69_M71	0.00%	-	-	714	46	351	158	86	462	625	684	704	711
M72_M75	0.00%	-	-	53	13	21	9	12	23	40	48	51	52
N77_N82	0.00%	-	-	523	42	270	156	90	297	438	493	512	519

[3] L68 is excluded as key sector because the effect is due to Imputed rent weight

Economic Activity	*Domestic Demand Change*			Output change	Change in Imports	Change in Value Added (GDP)	Change in Household Income	Employment change (Number)	*Cumulative Output*				
	%	Value	Total Change						First Round	Second Round	Third Round	Fourth Round	Fifth Round
O84	0.00%	-	-	13	3	9	7	10	6	10	12	12	13
P85	0.00%	-	-	16	1	13	9	13	6	12	14	15	16
Q86_Q88	0.00%	-	-	2	0	2	1	1	1	2	2	2	2
R90_R93	0.00%	-	-	0	0	0	0	0	0	0	0	0	0
S94_98	0.00%	-	-	3	0	1	1	1	0	2	3	3	3
Total impact	**0.05**	**-**	**14,970**	**25,934**	**4,057**	**9,222**	**2,731**	**4,639**	**21,689**	**24,456**	**25,431**	**25,765**	**25,877**

Source: Author's calculations based on INSTAT data.

BIBLIOGRAPHY

Australian Bureau of Statistics (ABS), Limitations of Input-Output Multipliers for Economic Impact_Assesment_(http://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/5209.0.55.001Main%20Features4201213?opendocument&tabname=Summary&prodno=5209.0.55.001&issue=2012-13&num=&view=).

Botric Valerija, (2013), Identifying Key Sectors in Croatian Economy Based on Input-Output Tables, Institute of Economics, Zagreb.

Chenery, H. B. and Watanabe, T. (1958) "International comparisons of the structure of production", Econometrica, vol. 26, 487-521.

Dietzenbacher, Erik, 1992, "The Measurement of Interindustry Linkages: Key Sectors in the Netherlands", Economic Modeling, 9(4), pp. 419-437.

European Commission, The European System of National and Regional Accounts (ESA 2010).

European Commission (2008), Eurostat Manual of Supply, Use and Input-Output Tables.

Gosh, A. (1968), Planning, Programming and Input-Output Models – Selected Papers on Indian Planning, Cambridge.

INSTAT (2016), Use and Input-Output Tables in Albania, http://www.instat.gov.al/en/themes/economy-and-finance/supply-use-and-input-output-tables/#tab2

José M. R-C; Frederik N. & Luis D. (2009), The adjustment capacity of the European economy examined with an input-output based key sector analysis:

towards a Review of the European Single Market, Seville, Spain.
McGilvray, J.W. (1977), Linkages, Key Sectors and Development Strategy, in: W. Leontief (ed.): Structure, System and Economic Policy, Cambridge 1977.

Miller, R. E. and Blair, P. D. (1985) "Input-Output Analysis, Foundations and Extensions", Prentice-Hall, Englewood Cliffs.

Oosterhaven, J. (2008): "A new approach to the selection of Key sectors: net forward and net backward linkages", paper of the International Input-Output Meeting on Managing the Environment: Input- Output & Environment, 9-11 July, Seville, Spain.

Rasmussen, P.N. (1956). "Studies in Intersectorial Relations, Amsterdam, North-

Holland P.C. Schultz S. (1977) Approaches to Identifying Key Sectors Empirically by Means of Input-Output Analysis", Journal of Development Studies, 14.

Reis H. and Rua A. (2006), An input-Output analysis, Linkages VS Leakages, Banco de Portugal, Lisboa, Portugal

Thage B. (2005), Symmetric input-output tables: compilation issues, paper presented at the 15th International conference on input-output techniques, Beijing.
Wixted, B., N. Yamano and C. Webb (2006), "Input-Output Analysis in an Increasingly Globalised World: Applications of OECD's Harmonised International Tables", OECD Science, Technology and Industry Working Papers, No. 2006/07, OECD Publishing, Paris.

EVOLUTION OF GIS TECHNOLOGY FROM ARCGIS DESKTOP TO WEBGIS AND ITS ADVANTAGES

LEDJO SEFERKOLLI, INSTITUTE OF STATISTICS
lseferkolli@instat.gov.al

NEXHMIJE LEÇINI, INSTITUTE OF STATISTICS
nlecini@instat.gov.al

MIRELA DEVA, INSTITUTE OF STATISTICS
mdeva@instat.gov.al

Abstract

Technology is advancing extremely, especially in recent decades, and the use of the geographic information system (GIS) has become increasingly important. GIS is a system that finds a lot of use in many areas, including statistics. The Institute of Statistics in Albania (INSTAT) has implemented for the first time GIS technology to conduct the Population and Housing Census (PHC), 2011. GIS technology supported the entire counting process in its three phases: the pre-Census, the Census, and the post-Census phase. The pre-Census phase started in 2009 with the implementation of a digital map infrastructure organized in GIS system, covering for the first time the entire territory of Albania at the building level. The demarcation of enumeration areas (EAs) was one of the critical activities of the Census phase, due to the lack of digital maps from the previous Census of 2001. For this purpose, a geospatial Census database was built. The delineation of enumeration areas and the printing of enumeration maps were based on this database. EAs were created in line with international standards and recommendations, where each enumeration area had an average of about 100 dwellings. In the census phase, GIS was used to support the fieldwork activities and data collection using EA maps. The use of EA maps in INSTAT supports not only major activities such as Census processes, but also a series of surveys planned by social statistics. Also, through the digital database-GIS, it is possible to publish statistical data through thematic maps, geospatial analysis, Census atlas, etc. That's why in the post-census phase, the maps were used for the representation, analyzing and the dissemination of the Census results. In this respect, the creation of a WebMap environment served as a new approach for INSTAT to publish Census 2011 data. The Geoportal [1]is fully operational and available to users on the INSTAT website, and thematic maps of Census 2011 are now available to interested users. These data are organized into three main thematic census groups such as: Population; Education and Employment; Housing and Living Conditions at different geographical levels. Using such an environment offers a new approach to users, as it allows them to manipulate data in their interest. Thanks to this portal, different users do not necessarily have to be cognitive to cartography to manipulate data. Also, the construction of this portal enables not only the mapping of census data, but also other statistical indicators produced by INSTAT, or from various administrative sources. Using such an environment offers a new approach to users, as it allows them to manipulate data in their interest.

KEYWORDS:

Census; Population; Enumeration Area; WebMap; WebGIS

1 http://instatgis.gov.al

1. INTRODUCTION

The Institute of Statistics in Albania has started preparations for the next census of population and housing. In the last census, INSTAT came up with two key innovations and one of them was the use of GIS technology. This publication shows how the GIS system is involved in all phases of census preparation activities and the advantages of using this system. But the implementation of GIS for statistical purposes has been very difficult because there was no geographic information in Albania before the 2011 Census. For this reason, INSTAT started to take the necessary steps to digitize all of this geographic information needed.

The process was organized in four main phases: a) map update in the field covering all urban areas in Albania, and selected rural areas; b) digitalization of all building boundaries and roads of Albania in INSTAT central offices; c) delineation of the entire territory of Albania into EAs for statistical data collection purposes; d) creation of a GIS Census database.
GIS was an important contributor during the enumeration phase of the PHC using EA maps, as in the orientation of field enumerators for identifying buildings, dwellings and households, as well as in monitoring the quality of their work and the logistical organization in fieldwork. In the third phase, after the census, GIS contributed to the analysis and presentation of the census results.
In INSTAT, the use of maps mainly consists in the presentation of statistical results in cartographic form. INSTAT's Geoportal was an innovation in the distribution of thematic maps on the internet and the growth of geo-statistics role. The thematic maps of Census 2011 data are available to users since May 2014, displaying different indicators at different geographic levels. With this new approach, INSTAT is supporting the needs of Albanian society and citizens, and has increased the transparency of data and communication with the users.
INSTAT also offers its support and contribute in projects dealing with geo-information, that help the production of national geospatial data. Albania is at the initial phase of building a NSDI with the objective to create a geodetic framework following European standards, the creation of geospatial data on national infrastructure, the design and development in the field of geo-information standards, and their implementation in institutions. Through the collaboration with ASIG (State Authority for Geospatial Information) [2] within the framework of the "National GIS" for the exchange of geospatial information, INSTAT provides for users the online map services (WebMapService / WMS) in their geoportal. Since INSTAT is responsible for the transformation and harmonization of its geographic data (two themes in the frame of the Albanian NSDI, population distribution and statistical units), according to INSPIRE[3] data specification, this institution is in the implementation phase for creating online map services (WMS) at the ASIG geoportal for "Population Distribution" and "Statistical Units" to contribute in the development of a national spatial data infrastructure, in line with EU standards and recommendations, and with a particular reference to the INSPIRE initiative.

2. WHY IMPLEMENTING A GIS DATABASE?

It has been very difficult to have a digital map of the census areas, in the previous censuses of the population and housing. During the 2001 Population and Housing Census fieldwork, interviewers have enumerated population using paper-based maps for urban areas of the country – here were included all major cities of Albania, while for rural areas interviewers were asked to go with pre-printed lists of household members. The old methodology used in 2001 has created problems during the various census stages and served as a clue point to start thinking about creating a new digital map - a new system - to be used in the Census of 2011. INSTAT implemented for the first time GIS to support the 2011 census process with the production of digital maps. Implementation of a new GIS database has shown to be very helpful due to its capacity to facilitate all operational census steps. The use of GIS database has supported the fieldwork activities in better management of staff and monitoring of the work. GIS has maximized data collection and processing, as well as production and distribution of data. All these advantages of GIS technology encouraged INSTAT to build and implement a GIS database to have a national infrastructure at the level of the buildings and roads that were useful for determining the boundaries of the enumeration areas. The GIS database was developed in two main phases: the first phase included field activities, which consisted mainly of collecting information on buildings and dwellings in Albania; and the second phase was work within INSTAT, where all the information collected by the operators on the fieldwork were digitized and used as the basis for creating the digital map of Albania.

[2] National Spatial Data Infrastructure

[3] https://inspire.ec.europa.eu

3. MAIN CENSUS PHASES AND GIS TECHNOLOGY

The implementation of GIS technology has been essential throughout all the census phases.

3.1 PRE-CENSUS PHASE

The first step in the process of creating a GIS database that was used to divide the entire territory of the country into small areas for statistical purposes (EA) was the map's update in the field. The map update has started in 2009, including all urban areas in Albania. At that period the old administrative and territorial division of the country was divided into 12 prefectures, 65 municipalities - which, according to the definitions in law, were considered mainly urban areas - and 308 communes - mainly consisted of rural Albania, but in rare cases within the communes were included also urban centres (cities). With the new administrative-territorial division of 2014, Albania is divided into 61 municipalities.

All the urban areas of Albania were updated in the field using 2007 satellite images. At the same time, INSTAT Central Offices began the work on the digitalization of all buildings identified in Albania during the pre-census fieldwork. The entire digitalization process was completed using the Esri ArcGIS software.

Digitalization was carried out in two steps:
a) In urban areas, where the digitalization was based on maps coming from fieldwork with the information about buildings and roads/streets.

b) In rural areas, digitizing was carried out without data collected on the fieldwork, the operators digitalized all information based on satellite images provided by the government of Albania. All this process was fully automated. The final result was the preparation of a database that served to create enumeration areas.

3.2 CENSUS PHASE

The Population and Housing Census 2011 was the first year Albania started using the digital maps for data collection on the fieldwork. The final result of the processes was the production of printed EA maps. Albania was divided approximately in 12,000 enumeration areas. For administration effects in the fieldwork during the enumeration phase, approximately 2300 controllers and 130 supervisor areas were created. In the census phase, the GIS supported activities and data collection on the field, using EA maps. Figure 1 shows an example of the EA map used by enumerators to collect statistical data.

3.3 POST-CENSUS PHASE

GIS has been very useful in the final census phase, in publishing data through production of statistical thematic maps. GIS was used in the post-census phase to analyze and disseminate statistical data through thematic maps with different indicators, as shown in Figure 2.
The first activities that were carried out in the post-census phase were mainly the control and publication of preliminary data, data entry, quality control, editing and analysis, as well as publications.

Figure 1: EA map example

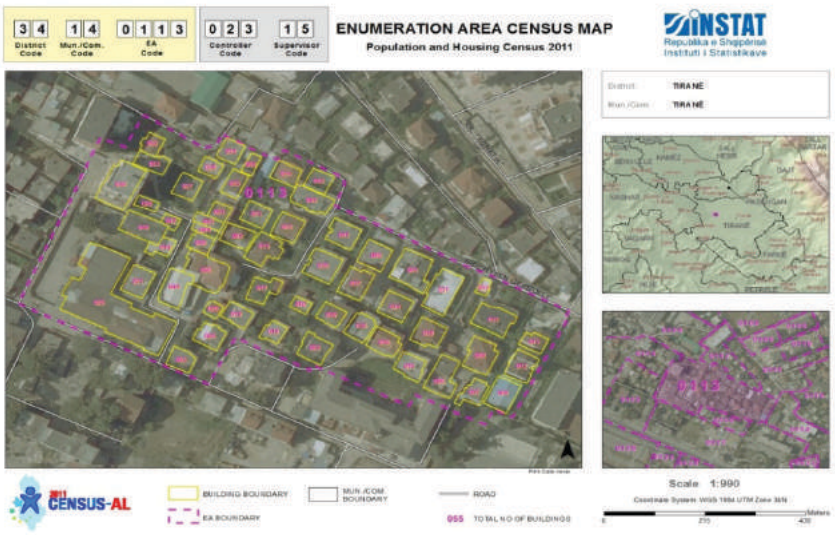
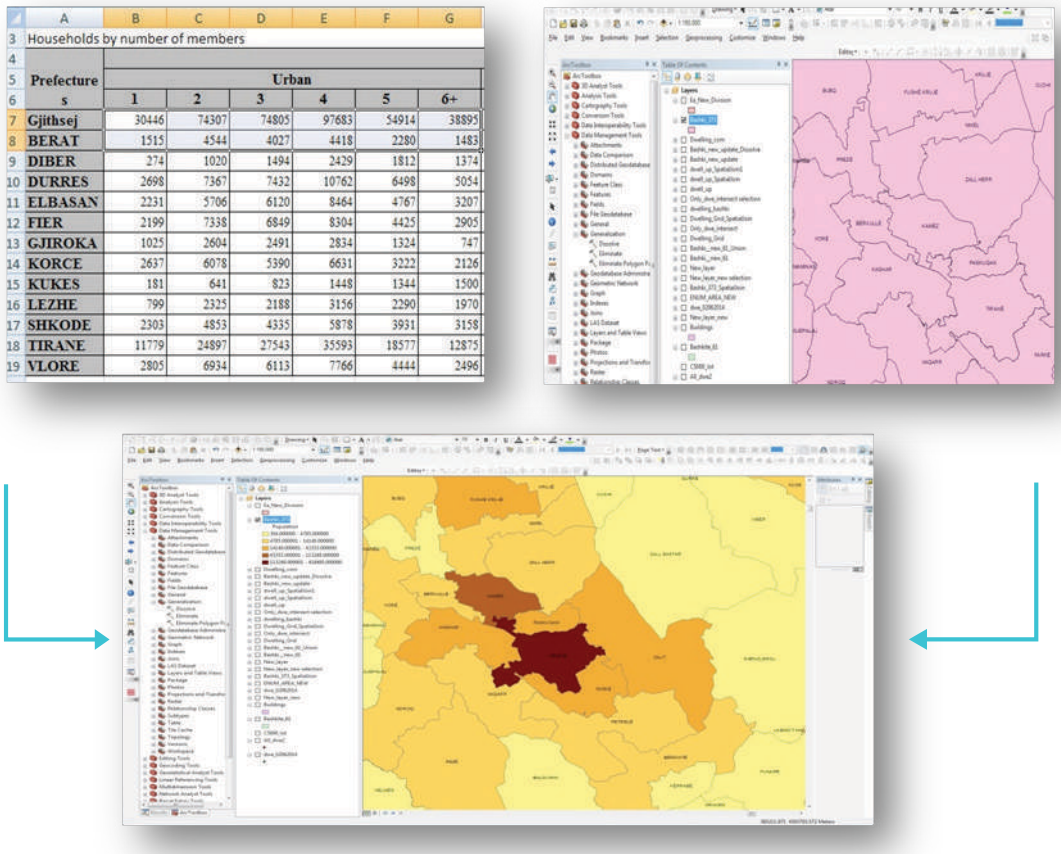


Figure 2: Example of a thematic map



Households by number of members						
Prefectures	Urban					
	1	2	3	4	5	6+
Gjithsej	30446	74307	74805	97683	54914	38895
BERAT	1515	4544	4027	4418	2280	1483
DIBER	274	1020	1494	2429	1812	1374
DURRES	2698	7367	7432	10762	6498	5054
ELBASAN	2231	5706	6120	8464	4767	3207
FIER	2199	7338	6849	8304	4425	2905
GJIROKA	1025	2604	2491	2834	1324	747
KORCE	2637	6078	5390	6631	3222	2126
KUKES	181	641	823	1448	1344	1500
LEZHE	799	2325	2188	3156	2290	1970
SHKODE	2303	4853	4335	5878	3931	3158
TIRANE	11779	24897	27543	35593	18577	12875
VLORE	2805	6934	6113	7766	4444	2496

4. WHY WEBGIS?

WebGIS is a kind of distribution system that is used to display and analyze geospatial data on the internet. WebGIS can be defined as any GIS that uses the internet to communicate between a server and a user. The difference between DesktopGIS and WebGIS is that the first one is used by trained and experienced GIS professionals. They use this software for data creation and processing, while WebGIS is more flexible and aims for a wider audience, including different users such as researchers, media, the public in general who can use GIS even without having knowledge on this field, without necessarily being a cartographer. WebGIS is designed for simplicity, convenience, making it much easier to use than GIS desktop. The web mapping application built by INSTAT was created precisely in function of this purpose; to bring closer users of all fields with statistics and geo-statistics. INSTAT built the WebGIS application to provide users the thematic maps with Census 2011 data, to make statistics visualization and geo-spatial data analysis faster, to facilitate the comparison of different indicators at different geographical levels, and to increase the transparency of data and communication with users. The portal[3] will serve to recognize geo-referenced Albania in terms of many economic and social characteristics at the prefecture, municipality/ commune and grid level. Currently, about 100 indicators from Census 2011 are published in the INSTAT geoportal for each geographical level, and seven indicators at 1km^2 grid level out of 13 indicators to be published after Census 2020 according to European standards and recommendations by Eurostat, and in line with the INSPIRE Directive.

[3] www.instatgis.gov.al

5. FROM ARCGIS DESKTOP TO WEBGIS

Nowadays with the spread of most advanced technology in communication, and the increased demand for statistical information, INSTAT is trying to be as close as possible to the user's need. Users from different fields want to have access to statistical information as quickly as possible. Possessing a mobile device allows them to have access in any time to the required information. Thus INSTAT tried to take the first steps to build a new way of communication with current users and potential users by launching a WebGIS portal to them.

Behind the WebGIS application lays a huge work done by the Cartography and GIS sector. Switching from GIS desktop to WebGIS passed through several phases. Initially, the optimized geodatabases for web services were created and prepared, such as preparation of indicators, import of indicators, quality control, etc. Four geodatabases were created to optimize the performance of web services:

• Grid.gdb for the indicators referred to the 1x1km INSPIRE grids
• Municipality.gdb for the indicators referred to the Communes and Municipalities
• Municipality.gdb for the indicators referred to the Municipalities[1]
• Prefeture.gdb for the indicators referred to the Prefectures

Each File Geodatabase will be the datasource of a specific web service built with a specific ArcMap project. ArcMap is the principal instrument to share a map as a web service through the web. Publication is done through ArcMap into ArcGIS for Server, to access later on ArcGIS Online which is a key component of WebGIS, to proceed then with the configuration of web maps such as: description of indicators, presentation and comparing of indicators in graphic form etc.
After the map with the data is ready, it is published in the INSTAT geoportal where users can access the statistical information they need.

The required statistical information comes not only classically through tables or px-axis integrated into INSTAT website, but it is also reflected on the maps. The WebGIS application can be accessed by various mobile devices. The application is dedicated to all users, professionals and non-professionals, and gives great benefits to users, allowing them to manipulate the data according to their needs.

4 With the new administrative-territorial division INSTAT in 2015 re-published the 2011 census data at the 61 municipal (Law No. 115/2014 dated 31.7.2014)

Figure 3: Atlas Web System Architecture

The 2011 Census data thematic maps are available to all users interested in statistics. Census data are shown based on census questionnaires and are classified into 3 main categories, which are: Population; Education and Employment; Housing and Living Conditions at different levels of data display.

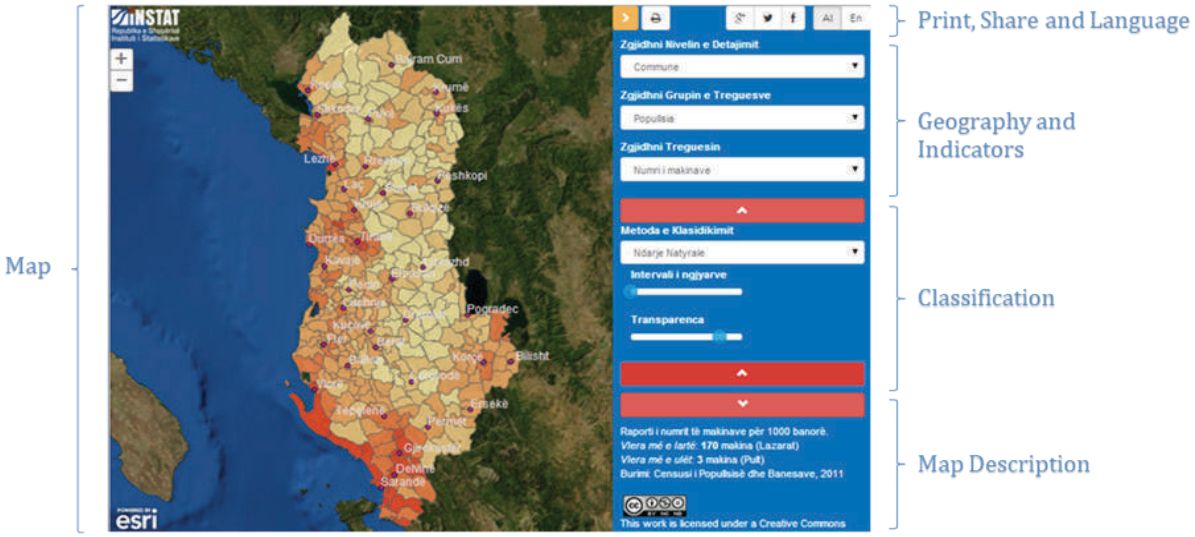
6. HOW WEBGIS WORKS? NAVIGATION AND USER INTERACTION

The WebGIS application structure is very easy to understand and use, available in two languages; Albanian and English. Through this geoportal, all interested users can easily access cartographic information as long as they have internet connection. Users can interact with the application by selecting statistical indicators, where they can create and print thematic maps of statistical indicators for different geographic levels. The web atlas has several GIS tools available to navigate to the portal, such as the tool pan which allows users to move through map, zoom in/zoom out etc. The main advantage of INSTAT's online application is that it allows users to individually manipulate data, ways of presenting maps, data classification, selecting the indicators for which they are interested, geographic level they need, selection of colours, etc.

Users can navigate through different panels:

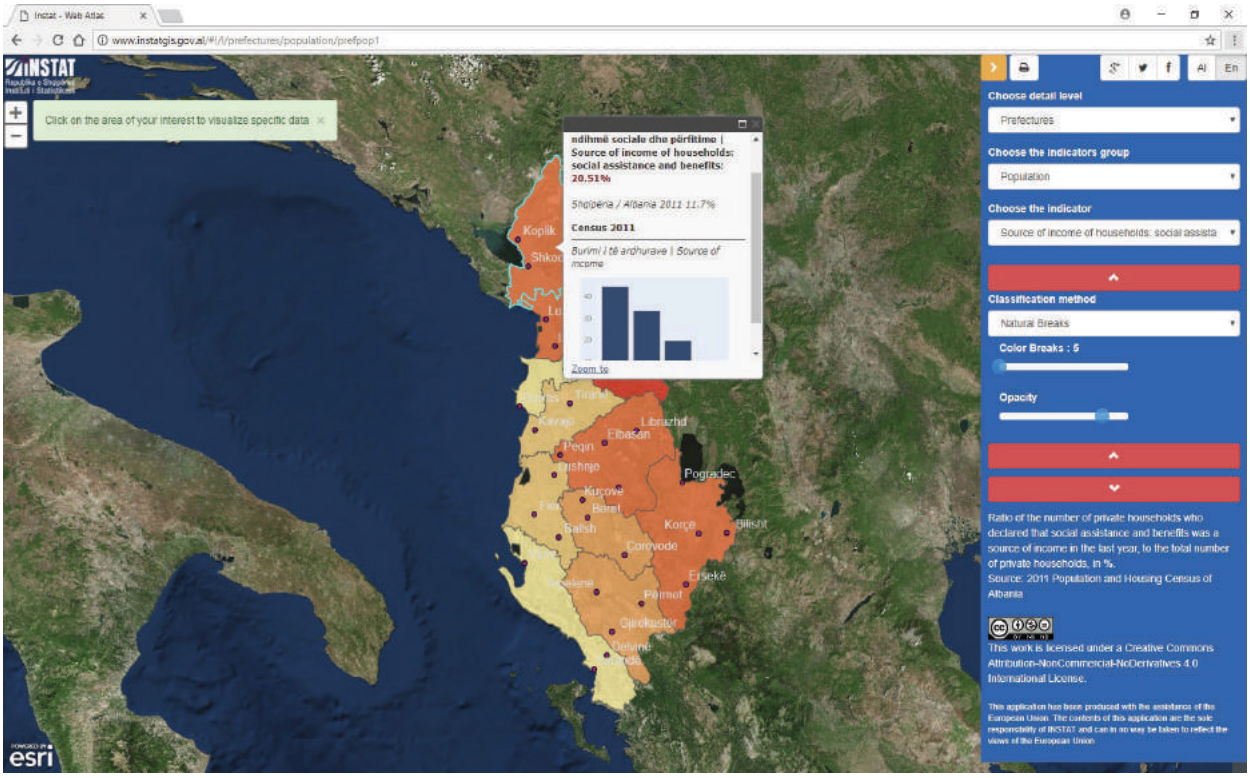
• Accessing the web geoportal, users will be displayed the main view of the "map" with the Albania border focused on it.
• The part of "geography and indicators", which allows users to choose one of the four levels, (Local Units, Prefectures, Grid 1km2), to map the indicator they want. Initially, users have the option by choosing one of the geographic levels to display the data. There are four levels in the respective panel: prefectures, municipalities / communes according to the old administrative division, 373 local unit[5], 61 municipalities[6] according to the new administrative division and the 1km2 grid system. In the indicators panel, users can choose one of the three major groups; the population, education and employment, as well as housing and living conditions.
• The "classification" method allows the user to manipulate and analyse the data in any way they want: the classification method (manual, natural-breaks, equal-interval, quantile, geometrical interval), the number of class breaks, the transparency of layer of thematic map, the colour ramps.
• The "map description" where users can find the description of the selected indicator.
• The "print, share and language", which allows the users to select the language (in Albanian or English language), to print the map, to share the map in different social networks; "google plus", "twitter", "facebook".

Figure 4: WebGIS interface



The maps can be printed from the website by the users according to the model created by INSTAT. Also the users can access these data via map services (WMS), from ArcGis software, QGIS and ArcGIS online. Users have the option to choose a specific geographic unit for a specific indicator on the map created by them, where through "pop-up" window they can read the information on the selected indicator.

Figure 5: Graphical presentation of data



Regarding to the application, it's being considered to improve some aspects of the web interface after CENSUS 2020; adding other statistical levels i.e. NUTS level; publication of CENSUS 2020 data at the grid level of 13 geocoded indicators, following the recommendations of the GEOSTAT project (EUROSTAT) and in accordance with the INSPIRE directive; creation of maps "Story map" etc., although in the future, the WebGIS application is also thought to be used for the integration of statistical and geospatial data and the creation of metadata, in line with the initiative undertaken by UN-GGIM[7] .

7. CONCLUSIONS AND FUTURE WORK

This paper aims to provide a general overview of the advantages by using GIS technology in census operations, from the pre-census stage, to the publication of data, and not only. Also, this technology is used to support other statistical developments belonging to INSTAT.
GIS is not just about where you are, but it's about exploring the world and seeing it from the others point of view. Jumping from ArcGIS desktop to WebGIS was another important development in presenting statistical information.
This new method of data presentation brings statistics even closer to users. In the future WebGIS will be enriched with information and other statistical indicators from different sectors; data from agriculture statistics, economic and social statistics.

INSTAT is always in touch with users from whom it receives feedback, and will try to meet their requirements in an efficient and responsible way towards statistical information.

[7] http://ggim.un.org/ggim_20171012/about.html

BIBLIOGRAHY

Albanian Official Gazette 137/2014

INSTAT 2012. 2011 Population and Housing Census of Albania, (Tirana 2012)

INSTAT 2014. Statistical Atlas of Albania, (Tirana 2014)

Law No. 115/2014 "On the administrative-territorial division of Local Government Units in the Republic of Albania".

DIFFERENCES BETWEEN SURVEY AND ADMINISTRATIVE DATA

CASE OF EMPLOYMENT AND WAGES INDEXES

REZARTA MYRTOLLARI, INSTITUTE OF STATISTICS
rmyrtollari@instat.gov.al

Introduction

In the production of official statistics, for a certain phenomenon, the available data may come from statistical surveys and administrative sources. Nowadays, the combination of these two sources is a promising and innovative strategy which affects the quality and quantity of research and increases the potential of data (Künn, 2015). However, this usage is often accompanied by serious challenges, given the simple fact that the purpose of designing the two sources is different. Administrative data are defined as data sets collected by government institutions or agencies for tax, benefit or public administration purposes (UNECE, 2011). On the other hand, data from surveys are collected specifically for statistical purposes. This article examines the differences between administrative sources and surveys focusing specifically on common variables of payroll as an administrative source and those from surveys conducted in INSTAT.

According to Penneck (2007) surveys differ from administrative data in the sense that they are specifically designed for analytical purposes, so coverage of population, definitions, methodology and time can be designed to meet these analytic needs. However, the sample size might be a problem if it is small since large-scale surveys are expensive and small-scale surveys have limited use. Samples are also subject to errors and non-response bias. In addition, mentions Penneck, we cannot be sure of the accuracy of business survey responses, compared, for example, with the administrative data collected for tax purposes. Administrative systems also require data from individuals, but the latter often see this as an indispensable part of the administrative process rather than as an additional statistical burden. The following sections will examine some of these issues in detail based on the work of Johnson and Moore (2008), illustrating them with concrete examples from the case of estimating average wage and employment indicators from two different sources in INSTAT.

1. POPULATION COVERAGE ISSUES

A system of administrative records defines the population covered by legislation based on the scope of the program intended for registration. This population is often limited by specific demographic or economic characteristics. According to Johnson and Moore (2008), in some cases individuals may need to undertake some actions to become part of the administrative system (e.g. registration of farmers in the tax and social security system by obtaining the NIPT to benefit from government support schemes). It is therefore important, say the authors, to consider what encourages individual units to be part of schemes. There may be some favouring factors for some individuals to avoid registration, especially if their circumstances place them close to a threshold that requires obligatory participation or gets associated with financial costs (p.10), such as setting a minimum wage on which contributions are calculated. Another factor is the change of policies that may fluctuate the population taken in study from year to year.
The Federal Committee on Statistical Methodology points out the differences on the unit of interest. The study unit needed for statistical purposes often focuses on the characteristics of groups formed by units (e.g. enterprises operating in a particular activity or large enterprises), while the administrative data focus on identifying specific units so that based on their individual characteristics (e.g. full-time employees or dual employment) certain actions can be undertaken. Thus, the differences in the entity reported in the tax statements limit the usefulness of the data for some types of research. Johnson and Moore explain that the target population of a survey is determined by the purpose of the study, the sampling frame availability, and the cost of the sample. Population for most surveys is derived from existing sources, such as population data based on geography, address lists, or other administrative sources. Often one of the most difficult issues in designing a survey is finding a suitable population (Lessler and Kalsbeek, 1992). If the population frame chosen for sampling is not suitable, it can lead to under coverage problems which may affect the results obtained from the survey data. Another problem arises if a survey targets a population that is difficult to find or measure. Directly related to the availability of a population is the potential cost of receiving population data and the cost of interviewing a sample of the desired size. For target populations that are difficult to find, simply the cost of increasing sample size to provide better coverage may be obstructive to undertake such an initiative (Johnson and Moore, 2008, p.13).

2. CONTENT ISSUES

Johnson and Moore list several content issues that need to be considered while working with administrative and survey data. One of them is the purpose for which administrative data are collected, which may have a significant impact on their usefulness for statistical purposes regarding the amount of available data, data definitions, consistency between different time periods and data quality. The authors argue that many times the usefulness of administrative registry systems is limited because only those variables needed to administer the tax and tax payment program are collected. These variables can only be a small part of the data reported in an administrative form (p.15). In addition, because program requirements are defined by legislation, the concepts and definitions of variables used to meet program needs do not necessarily match those required for social or economic analysis (Brackstone, 1987). For example, one of the problems faced in our administrative data comes as a result of using the concept of working days instead of working hours in the declaration of the taxpayer's payroll. This difference makes it difficult to compare employment data with those of national accounts. Such factors may pose serious limitations on the overall usefulness of the administrative data systems or require that the administrative agency undertake collecting and / or editing additional data, causing financial costs and delaying the availability of data.

An important aspect of the data content is continuity over time of the variables included and their definitions. Coverage and content in administrative data systems may be subject to discontinuities resulting from changes in laws, regulations, administrative practices or the scope of the program (Brackstone, 1987). For example, the revisions of the law on the minimum and maximum wage level made that the minimum monthly base salary for employees required by any legal or natural person, domestic or foreign, is 24.000 ALL from 22.000 ALL that was before this period. Such changes have a significant impact on the statistical uses of data for comparisons between periods.

Administrative data systems also can not ensure perfect data quality. Information that might be important for statisticians, but less important for administrative purposes, is often reported and processed imperfectly, noticed Johnson and Moore (2008). Here we can mention the choice of profession by the person who makes the declaration of salary and wages. The variable that indicates the profession category has a secondary importance for the administrative agencies as long as the person

regularly declares his or her contributions. On the other hand, this variable is of particular importance for the production of statistics about average wage by group of professions. The other variables used mainly as secondary or complementary information may be of low or even incomplete quality (as in the case of working days for which the declaration is usually a standard of 21 or 22 business days). This phenomenon may also occur with data specifically collected for statistical purposes using existing administrative channels such as in the case of enterprise activity classification in Nace Rev.2 collected by the administrative agencies for the account of INSTAT. These variables may be of lower quality if their priority is not too high for the administering authority or the entity providing the information (Jensen, 1987). Another issue pointed out by Johnson and Moore (2008) is data reliability which can be affected if the information provided to the tax entity can cause profits or losses for the declaring subject. Moreover, given that the data collected and processed for administrative purposes are generally given priority over what is required for statistical purposes, the amount of processing required to provide administrative data suitable for statistical purposes may affect the time that these data are made available to statisticians, argue the authors.

Many of the issues raised above are best addressed and resolved through surveys, stated Johnson and Moore. However, the authors notice that other content and validation issues of some kind appear in the survey data. The key issue here is the voluntary nature of responses to surveys versus the legal obligation to participate in administrative data programs. The respondent needs to be persuaded to give their time and the information required despite the fact that there are no consequences if he or she refuses and there are no benefits if the survey is filled. Still, if a respondent agrees to participate in the survey, it is possible that he or she refuses to answer the questions in a "real" manner (p.21).

For respondents who agree to attend and respond to all survey questions, the measurement error is still a concern for the survey data, state Johnson and Moore (2008). Respondents can "think" answers to questions or they may have difficulty to remember past events. Other typical measurement errors include rounding of amounts, misunderstanding of questions and changing responses due to fears about the disclosed data or the desire to protect privacy. Numerous studies exist regarding error measurements and their effects on observation data (Lessler and Kalsbeek, 1992). While it is true that, for administrative data, non-response is not an essential problem, it is not clear whether administrative records are always more accurate than observation data, report the authors. An example would be the number of employees declared by the enterprise; some companies intentionally may declare a lower number of employees into their statements to reduce their tax obligations. The same individuals can report the true value in responding to a questionnaire as there are no legal consequences if true value is stated. Another content issue for the survey data is the timeline of data. While many simple surveys are carried out at frequent frequencies, monthly or quarterly, most of the most complex surveys occur annually or even rarer. Costs and other resource constraints are major factors in timely use of survey data. A final content issue for surveys elaborated by Johnson and Moore (2008) is validation of data. According to the authors sometimes it is possible to conduct validation studies after a survey has been completed and these studies add additional costs to the survey or validation of selected data variables can be carried out with external sources such as censuses or administrative records, but there is often no validation source (p.25).

3. PRIVACY ISSUES

In their work Johnson and Moore (2008) consider data privacy as a very important issue. The authors explain that any use of administrative data for research purposes should take into account the laws that protect the privacy of the data. The research of administrative data is often limited to uses within the scope of an agency mission and should be carried out only by persons working for the agency as employees, contractors or under Memorandum of Understanding that allow employees of various institutions to exchange the data. The way the public percept privacy protection of their data has a direct impact on the continuity of the levels of declarations (p.26). Often, because of these factors, the available data does not contain identifying variables. For example, in the case of individual data from the administrative source, variables which directly identify the subject are missing. Of course in another scenario the availability of these variables could lead to wider statistical use and a combination of data from different sources.

However, the authors emphasize that data confidentiality is of great importance to the current and future success of any administrative observation and registration. If the subjects do not believe that their data is sufficiently protected, response rates and overall data quality will be subject to deformation. Confidentially and privacy laws offer significant protection against potential abuse of data (Johnson and Moore, 2008).

4. EMPLOYMENT AND WAGES AND SALARY DATA

Both administrative sources and surveys (such as Quarterly Short-term Statistics Survey) provide important information at quarterly frequency regarding the number of employees and salary fund. This data is used to calculate wages and salary index and the employment index. Administrative records have richer demographic information about the individual and detailed data on social and health contributions. On the other hand, survey data is more limited, including only the number of employees and the wages and salary fund of the surveyed enterprise.

The most important changes between the two sources, as theoretically are discussed above, relate to the survey unit, population coverage, and sample size. STS is a quarterly survey where the surveyed unit is the enterprise and the main variables required are Net Sales, Industrial Production, Construction Production, Average Number of Employees, Wages and Salaries Fund, Production Prices, Import Prices, Construction Costs (INSTAT, 2017). All produced indicators are expressed in indices, in annual and quarterly changes. The study unit for the administrative source is always the individual, and the average salary indicators are expressed as absolute values. The size of the STS sample is limited due to frequency and cost, and the survey does not cover all economic activities, leaving out the assessment of the agricultural activities (section A), those financial services and insurance (section K), real estate (section L), public administration (section O), education and health (sections P & C), as well as arts, entertainment and entertainment activities, other services and activities of international organizations (i.e. sections R, S, T, U) which are outside its coverage area. This means that the quarterly information from the survey about employment and salaries is missing for these industries. On the other hand, the information from the administrative source includes individuals and enterprises in all economic activities.
The change in methodology has a direct impact on the estimates derived from each source. In addition, STS estimates are not particularly focused on estimating average wages and the lack of detailed employee information (e.g. full-time or part-time employment, dual employment, contribution category etc.) makes it is impossible to apply the same methodology as the one used to estimate the average salary from the administrative source. Furthermore, the data from surveys are subject to the weighing process, while the assessment from the administrative source is straightforward.

5. CONCLUSIONS

Nowadays there is a need to satisfy a growing demand from users for good quality statistics, enabling faster measurement of new phenomena. At the same time, the demands of these users are in line with the needs of today environment that the burden placed on businesses and citizens diminishes (Laux, Baigorri, & Radermacher, 2009). Therefore, the use and combination of administrative or secondary data by statisticians is seen as a necessity in the present days, but it is also accompanied by a number of challenges. Some indicators, such as those discussed above, can be produced with data that can be derived from both administrative and statistical sources, but the fundamental structural differences between these two sources, as well as changes in the applied methodology, result in differences in estimations and, of course, the final results obtained from them. These changes are present in almost all dimensions of quality, such as relevance, accuracy, timeliness, accessibility, comparability, and timing. For this reason, users should be aware of these changes at the time of using the estimations from different sources and should understand the origin of data, their collection and use, in order to avoid mistakes and misunderstandings. This allows them to choose the indicators that fit the best their study goals (Laux, Baigorri, & Radermacher, 2009). More than just competing sources, administrative data and surveys should be seen as complementary sources. As Kapteyn & Ypma (2007) say, the question of whether administrative resources or observations show "truth" is almost a philosophical question.

BIBLIOGRAPHY

Brackstone. (1987). Statistical uses of Administrative Data: Issues and Challenges. Statistical Uses of Administrative Data Proceedings, (fv. 5-26).

INSTAT. (2017). Statistikat Afatshkurtra- Raporti i cilësisë. Tirana, Albania: INSTAT.
Jensen, P. (1987). The Quality of Administrative Data From a Statistical Point of View, Some Danish Experience and Considerations. Statistical Uses of Administrative Data Proceedings, (fv. 291-300).

Johnson, B., & Moore, K. (2008). Comparing Administrative and Survey Data. IRS Statistics of Income Working Paper Series.

Kapteyn, & Ypma. (2007). Measurement error and misclassification: A comparison of survey and administrative data. Journal of Labor Economics, 513-551.

Künn, S. (2015). The challenges of linking survey and administrative data. IZA World of Labour, 214.

Laux, R., Baigorri, A., & Radermacher, W. (2009). Building Confidence in the Use of Administrative Data for Statistical Purposes ., (f. 9).

Lessler, & Kalsbeek. (1992). Nonsampling Error in Surveys. New York : John Wiley & Sons.

Penneck, S. (2007). Using Administrative Data for Statistical Purposes . ICES-III. Montreal, Quebec, Canada.

UNECE. (2011). Using Administrative and Secondary Sources for Official Statistics - A Handbook of Principles and Practices, United Nations Economic Commission for Europe. New York and Geneva: UNITED NATIONS.

METHODOLOGICAL AND THEORETICAL CONCEPTS OF POVERTY IN ALBANIA

LEDIA THOMO, INSTITUTE OF STATISTICS
lthomo@instat.gov.al

TERANDA JAHJA, INSTITUTE OF STATISTICS
TJAHJA@INSTAT.GOV.AL

RUZHDIE BICI, INSTITUTE OF STATISTICS
rbici@instat.gov.al

BLERTA MUJA, INSTITUTE OF STATISTICS
bmuja@instat.gov.al

Abstract

Living standards, poverty and wellbeing, were measured until 2012 using the Living Standards Measurement Survey (LSMS). In 2016, INSTAT presented a new method of calculating these indicators. Until 2012, poverty measurement was based on determining the absolute poverty line based on consumption expenditures. From 2016-ongoing, poverty will be measured based on the relative line and the equivalent household income.

The purpose of this article is to inform about methodological changes occurred. Taking into account the methodological changes, the information and results is expected to be different; the range of indicators will be wider, focusing on monetary and non-monetary estimates of poverty and comparable to other EU-countries. The analysis is focused on narrative and informative explanations for the new methodology involved in INSTAT activities.

KEY WORDS:

Income; Living standard; Poverty; Relative poverty line

1. INTRODUCTION

The Living Standard Measurement Survey (LSMS) is the only source of information to measure the living standard, poverty, and wellbeing of Albanian household until 2012.
This survey collects a series of monetary and non-monetary indicators, bringing a variety of information to different users, and provides a necessary tool for policy maker and strategies.
LSMS was conducted for the first time in 2002, followed by two other surveys every three years, respectively in 2005, 2008 and 2012 (INSTAT, 2013).
Through LSMS it is tended to measure poverty by consumption expenditures and using the absolute poverty line. LSMS data are widely used by different users inside and outside the country.
From year 2016 in the activities of INSTAT was included the Income and Living Conditions Survey (SILC). SILC measures poverty based on total income of individuals or households and uses the relative poverty line. The SILC methodology is based on the Eurostat methodology and is comparable with EU countries (Eurostat, 2017).

2. METHODOLOGY OF CALCULATING POVERTY

2.1 METHODOLOGY OF CALCULATING MONETARY POVERTY BASED BY HOUSEHOLD CONSUMPTION EXPENDITURES.

For the first time, the poverty based on the consumption expenditures was carried out by the LSMS in 2002. The data of this survey were representative at four regions and by urban and rural area[1]. Other LSMS surveys conducted in 2005 and 2008 followed the same methodology. In 2012, in order to have more representative (from 4 Regions to 12 Prefectures), was increased the number of selected households in the survey. The results of this survey give a clear picture of expenditures, living conditions and poverty.

The LSMS calculates poverty by taking in consideration the necessary expenditure to provide food products and an expenditure level of non-food basic necessities. In this way is calculated the absolute poverty line. For the calculation of the monetary poverty line it is used the methodology of "Cost of Basic Needs" (Ravallion and Bidani, 1994). The poverty level is calculated using the food basket consumed by the individuals in the second to the fourth lowest deciles.

Taking into consideration the FAO[2] recommendations on the minimum calorie requirements according to age and sex, and adjusting these to the population distribution in Albania in 2001 (INSTAT, Census, 2001), it was estimated that the per capita required calorie intake was set at 2,288 calories per day.

The non-food component of the poverty line was calculated, taking into consideration the percentage of non-food expenditure of those households that spend for food consumption an amount approximately equivalent to the food poverty line. The food poverty line or extreme poverty line was set at 3,047 ALL per month, whereas the poverty line has been set at 4,891 ALL per month at constant prices (2002).

All the calculations done in 2005, 2008 and 2012 surveys have as base line the year 2002. For this reason, all results are deflated to bring the real values, which mean that they are cleaned from the influence of the price changes in the respective periods.

To calculate the total poverty line, first is calculated a food poverty line, or the cost of obtaining a certain minimum amount of calories, and then augments it by making an allowance for non-food basic necessities. The non-food component is calculated as the average of non-food share expenditures of households, assuming that they spent for food roughly the same amount as the food poverty line.

2.2 METHODOLOGY OF CALCULATING MONETARY POVERTY BASED ON HOUSEHOLD INCOME

In 2016, for the first time in our country was conducted Income and Living Conditions Survey. The main purpose of this survey is to study the living conditions of households mainly related to their incomes both at national and European level. This survey is a basic source for generating comparable statistics on income distribution and social exclusion at European level. The EU-SILC survey is implemented in all EU member states on a regular annual basis. Income, social inclusion and living conditions statistics cover the objective and subjective aspects of these topics in both monetary and non-monetary aspects for households and individuals. The expected results will provide comparable statistics to other countries and will

[1] http://instat.gov.al/en/themes/social-condition/living-standard-measurement-survey/#tab4

[2] Food Agriculture Organization

help the country with reliable data in the area of poverty, social exclusion, and household income, helping in anti-poverty policies.

EU-SILC survey provides two types of annual data:
- Cross-sectional data pertaining to a given time or a certain time period with variables on income, poverty, social exclusion and living conditions
- Longitudinal data pertaining to individual-level changes over time, observed periodically over a four-year period.

Since the SILC survey contains panel data, the sample each year consists of 4 rotational groups, which have been in the survey for 1-4 years. Any rotation group remains in the survey for 4 years; every year one of the 4 rotational groups from the previous year is dropped and a new one is added. According to the methodology for measuring poverty, the poverty line is calculated with its relative concept (poor in relation to others) and is defined at 60% of the median of total equivalised disposable income of the household (disposable income per person), using the modified OECD equivalised scale. Total equivalised disposable household income is considered the household net income (that is, income after taxes, social and health contributions) from all household members.

Through SILC, equivalised disposable income of the individual is considered the total disposable income of the household after being divided using the modified OECD equivalised scale.

It is pointed out that in the income distribution per person it is suggested that each member of the household possesses the same income, i.e. the equivalised disposable income of the household. This means that each member of the household has the same standard of living. Consequently, in the income distribution per person, the income that is attributed to each person does not represent wages but, actually, an indicator of living conditions. All income values are calculated annually and are aggregated at household level.

The SILC methodology is complex and takes into account the same form of calculation as referred to Eurostat guideline (Eurostat, 2017). However, adjustments and changes are realized regarding the specifics of the country conducting the survey. This relates to the calculation of gross and net income where the taxation and insurance system changes, or other specifics related to transfers, social payments and pensions.
Equivalised disposable income is calculated as the total disposable income of the household divided by its equivalent size.

Equivalent household size refers to the OECD (OECD, 2008) modified scale and is calculated taking into account the number of adults in household (14 years and over) and children under 14 years old. The first adult of the household, 14 years old and over is counted as 1.0 person; other persons aged 14 or over are counted 0.5 persons and every household member who is 13 years old or younger is counted as 0.3 persons.

The function for the weight in the household size is:

W_{OECD}=1A+0.5OA+0.3CH

The income of a household with two adults and two children under 14 years old is divided with a weight of 1 + 0.5 + (2*0.3) = 2.1; of a household with two adults is divided with 1 + 0.5 = 1.5; of a household with two adults and two children above 14 years old is divided with (1 + 0.5 + (2*0.5) = 2.5, etc. Disposable income of the household is divided by the equivalent size of the household, obtaining equivalised disposable income per person and ranking from the lowest value to the highest value.

From EU-SILC the poverty line is calculated by taking into account what is known as the relative poverty line widely used as the official poverty line for measuring poverty in European countries, 60 percent of the median of the equivalised disposable income per person.

In this way, people who fall below the poverty line are defined as individuals whose equivalised disposable income is below the poverty line. This is also the monetary poverty line.

3. COMPARABILITY BETWEEN SILC AND LSMS

There are a number of differences and methodological differences in the two surveys that normal bring different expectations in variables and indicators, as shown in Table 1.

Table 1: Comparability SILC vs. LSMS

	LSMS	SILC
Sample	Representative in 4 Regions and the last survey at 12 Prefectures	Representative in 12 Prefecture and 3 Regions
	No panel survey	Panel survey
	Cross sectional	Cross sectional and longitudinal
Questionnaire	Household questionnaire and individual questionnaire for all members (except employment and communication section)	Household questionnaire and Individual questionnaire for all household members 16 year and over
	Standard and comparable in years.	Standard for core variables. Comparability of variables and individuals. A new module every year
Monetary Poverty	Absolute poverty line based in household consumptions	Relative poverty line based on equivalised income
Non- monetary Poverty	Unmet Basic Needs	Material Deprivation
Periodicity	Every three years	Every year
Comparability	Not comparable with EU countries	Comparable with EU countries

4. CONCLUSION

There are significant changes in the design, sampling, calculation methodology, and indicators produced by both sources. The SILC survey will be an innovation and a wide information source, enabling comparability over the years, as well as with European countries.
The two sources of information do not have comparability because through LSMS is calculated the absolute poverty line and through SILC is calculated the relative poverty line. Both surveys are multidimensional for the variety of indicators and for the information they collect.

The purpose of both surveys is to calculate wellbeing, living standard and poverty. However, different users using these two sources should consider that there is no comparative basis between them. Changes in methodology are essential and produced indicators, such as absolute and relative poverty, are two different concepts, although both measure poverty.

BIBLIOGRAPHY

Eurostat (2017), EUROPEAN COMMISSION, Directorate F: Social Statistics, Unit F-4: Quality of life DocSILC065 (2017 operation), Methodological guidelines and description of EU-SILC target variables, 2017 operation (Version may 2017);

http://instat.gov.al/en/themes/social-condition/living-standard-measurement-survey/#tab4.

INSTAT, (2013), Albania: Trends in Poverty 2002-2005-2008-2012, http://www.instat.gov.al/media/1312/living_standart_measurement_survey__2012_revised.pdf

OECD (2008), "Growing Unequal? Income Distribution and Poverty in OECD Countries (2008)", OECD, Paris.
Ravallion M. and Bidani B. (1994), "How Robust Is a Poverty Profile?", The World Bank Economic Review;

WHY IS GROSS DOMESTIC PRODUCT REVISED?!

ERJOLA GJIKA, INSTITUTE OF STATISTICS
ekarpuzi@instat.gov.al

ERJOLA ISMALAJ, INSTITUTE OF STATISTICS
eismalaj@instat.gov.al

ENI CELO, INSTITUTE OF STATISTICS
ecelo@instat.gov.al

MARINELA NUSHI, INSTITUTE OF STATISTICS
mnushi@instat.gov.al

Introduction

INSTAT's vision is to produce reliable statistics. The purpose is transparency regarding the methods used in compilation of statistics and the use of all available information at a given time. In order to produce the best estimation, the statisticians have to trade off between accuracy and timelines. Deadlines set for publication dictate the need for the review of statistics.

These revisions aim the improvement in the quality of previously published data. They are important because the information made available after the first publication suggests another picture of the reality which needs to be taken into consideration. Users do not always welcome small statistical changes after a long period of publication. For this reason, it is necessary for revisions to be based on what is known as "Revision policy" which provides a set of guidelines and principles, which promote transparency of processes and ensure trust and effective communication with users. This article analyzes the effects of the regular revisions of GDP by focusing on the compilation of GDP by Expenditure Approach.

1. CONCEPTS IN NATIONAL ACCOUNTS

National Accounts in Albania are built on a series of surveys and administrative resources that measure activity in the economy in different ways. Gross Domestic Product (GDP) in our country is measured in two ways according to:
- The production approach
- The expenditure approach

2.1 PRODUCTION APPROACH

The main indicators of the production method are output, intermediate consumption and value added. Production shows the value of goods and services produced in the economy over a year, while intermediate consumption is the value of products or services transformed or fully consumed during the production process. By subtracting intermediate consumption from production, we reach what is known as Gross Value Added (GVA).
Gross Domestic Product by production method is equal to gross value added by adding taxes and subtracting subsidies on products.

2.2 EXPENDITURE APPROACH

The expenditure approach brings the inverse panorama, estimating GDP according to the final internal uses of products and services. According to this method, the Gross Domestic Product is estimated as the value of final expenditures by consumers, non-profit institutions and the government, adding the gross capital formation and exports of goods and services, and subtracting imports of goods and services.
Regular revision of GDP on an annual basis is the result of updating and fulfilling the existing data sources with the latest information for a given year.

2. REVISION PERIODS

As in developed countries, where the GDP estimation is revised several times, in Albania this estimation is done in two phases:
a) The first stage includes semi-final evaluations of the year "t" which are performed in the period t + 15 months
b) The second phase includes the final estimates of GDP of the year (t). At this stage, the data of year (t), are revised once more when the data set for year (t + 1) becomes available. This is done in order to include any changes made to the year (t) and ensure consistency between two consecutive years.

These revisions are intended to describe more accurately the country's economy. It is useful to keep in mind the frameworks and timelines for the publication of continually improved estimations.

3. REASONS FOR REVISION

The public wants accurate data as soon as possible. To meet this need, INSTAT publishes semi-final estimations. Although these data are incomplete, they provide an accurate overall picture of economic activity. They give direction and trends of various components of the economy, providing valuable information to businesses and policymakers.

Changes in the semi-final and final estimations of GDP come as a result of:
• Addition of the number of financial statements of the year (t) integrated into the databases after closing the semi-final estimate
• A higher number of businesses that present the financial statements of year t + 1, reporting also the previous year
• Revision of foreign trade data such as imports and exports of goods
• Updating data from balance of payments
• Revision of data from the Agriculture Directorate

4. EFFECTS OF GDP REVISION

The following article shows the effect of the revision in GDP, in terms of key indicators of the expenditure method. For this purpose, the relevant values in the semi-final assessment of 2015, versus the final one, have been analyzed.

As a result of the revaluation, PPB at current prices changed from 1,427,799 million ALL on the semi-final estimate to 1,434,307 million ALL on the final one. Following the revision of the results, the real growth rate decreased by 0.01 percentage point.

Contribution of relevant components to real GDP growth has changed from the semi-final to the final one of 2015. In the semi-final estimation, HFCE contributed by 0.80% to GDP growth and GFCF with 0.97%. Meanwhile, in the final estimation, HFCE contributed with 0.73% and GFCF with 0.85%.

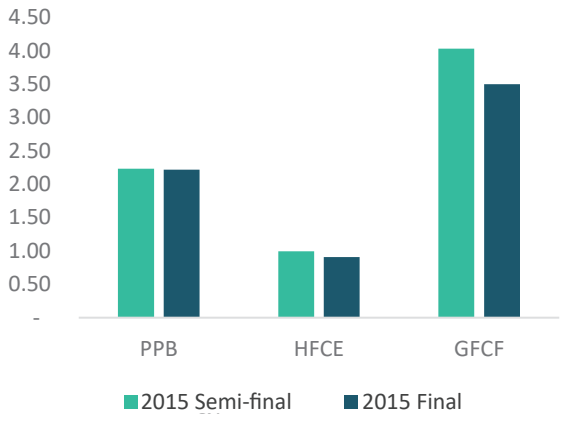
Household Final Consumption Expenditure (HFCE) consists of the expenditure incurred by households on individual consumption of goods and services, including those sold at prices that are not economically significant. It also includes various kinds of imputed expenditure of which the imputed rent is generally the most important one. Household final consumption expenditure account 79.9% of

gross domestic product in 2015 (according to the final estimation). In the semi-final estimation of 2015, the group of "Hotels, cafes and restaurants" has the highest growth rate in real terms, by 12.09%. In the final estimation, this group continues to have the highest growth, by 13.02% in real terms.

On the other hand, the highest decrease, according to the semi-final estimation, is found in the "Entertainment and culture" group by 12.39% in real terms. In the final estimation, this group has again the highest decrease by 11.53% in real terms.

In addition, the effect of the Gross Fixed Capital Formation (GFCF) is shown, which consists of the costs incurred for new capital or other specific expenditures carried out by the producer on the items or services in order to maintain at the same level, increase or expand their production capacities, or create new processing conditions in the future. As one of the main components of GDP in the expenditure method, gross fixed capital formation has a weight of 24.4% in GDP (according to the final estimate of 2015). Thus, the highest growth in GFCF, in the semi-final estimation of 2015, is in the group "machinery and equipment" with 15.85% in real terms. Also in the final estimation of this year, the highest increase in GFCF is in the same group, but at lower levels, by 15.81% in real terms. "Construction" also grew less, according to the final estimation compared to the semi-final, moving from 6.03% to 5.68% increase in real terms.

Figure 1: Real growth after revisions



5. RESULTS OF REGULAR REVISIONS

Changes between final and semi-final evaluations were observed, but are relatively small:

- Real GDP growth after revisions decreased by 0.01 percentage point.
- HFCE contribution to real GDP growth decreased by 0.07 percentage point.
- After the revision of data, the growth rate of the group "Hotels, cafes and restaurants" increased by 0.93 percentage point.
- The decrease rate of the group "Entertainment and culture" was softened, increasing by 0.86 percentage point as a result of the routine revision.
- GFCF contribution to real GDP growth decreased by 0.12 percentage point.
- The largest change resulting from the revision to GFCF, was at "Construction" group by 0.35 percentage point decrease in real terms, followed by "Machinery and Equipment" also falling by 0.04 percentage point in real terms.

Although revisions of GDP estimates are being made, as more complete data is provided, studies show that the overall picture of the economy does not change.

Measuring GDP is a work that is always in progress. It often takes months, or even years, to provide more complete and accurate information. Our semi-finals estimate has a good balance between accuracy and time, given the available data at that time. Successive revisions reflect the commitment of INSTAT to include the most complete resource when they become available and to improve methods for measuring a constantly changing economy.

6. PRACTICE OF OTHER COUNTRIES

Even in other countries, the main objective of revising statistical data is to improve the quality of information previously provided. There are various reasons for conducting common statistical reviews, where the main is to include new data related to the past, which could not be included at the time of compiling and publishing the first version.

The case of Germany

In order to be able to publish as early as possible the most recent data on "economic growth", the results are calculated first on the basis of incomplete and partially evaluated data. These preliminary gross domestic product results will be updated continuously in future releases when new statistical data become available. Preliminary results adjustments over time follow a well-known and published revision cycle. Including this new information is particularly necessary if there are breaks in time series. This is necessary for comparing data with the previous year. Once a year, a major national account review is carried out with a comprehensive review of previous years, which includes the last four years. These results are always published in August. In particular, this revision takes into account data from tax return statistics and structural business survey. Structural business survey is based, in particular, on the annual financial statements of enterprises and above all provides information on their costs. By contrast, preliminary calculations are based only on current output indicators.

One of the key challenges for achieving early economic growth estimations are turning points in the economy. GDP revisions have greater impacts on these cases, as the preliminary results of many initial statistics fail to catch these moments of change. In addition, there are special effects such as VAT increases or extreme weather conditions, the impact of which is difficult to evaluate. In this context, it is obvious that revisions may be more frequent and larger in a dynamic economy.

BIBLIOGRAPHY

"Revisions and the Income and Expenditure Accounts", Catalogue no. 13-604-M — no. 068, Statistics Canada

"Volkswirtschaftliche Gesamtrechnungen", Revisionsbedarf des Bruttoinlandsprodukts, Statistisches Bundesamt

Ahmad, Nadim, Sophie Bournot and Francette Koechlin. 2004. "Revisions to quarterly GDP estimates: A comparative analysis for seven large OECD country" OECD

Fixler, Dennis J. and Bruce T. Grimm. 2003. "Revisions, Rationality, and Turning Points in GDP". U.S. Bureau of Economic Analysis.
Graeme Walker, Andrew Walton and Tiffany Georghiades, "Why is GDP revised?",Office for National Statistics, UK

http://www.instat.gov.al/al/temat/ekonomi-dhe-financë/llogaritë-kombëtare-gdp/publikimet/2018/produkti-i-brendshëm-bruto-final-2015-dhe-gjysmë-finale-2016/

McKenzie Richard, Elena Tosetto (OECD) and Dennis Fixler (United States Bureau of Economic Analysis) 2008. "Assessing the efficiency of early release estimates of economic statistics". OECD.

Peter van de Ven and George van Leeuwen, "Discussion", Vol. 20, No. 4, 2004, pp. 607–614,Journal of Official Statistics

Smith, Philip.1977. "An Analysis of the Revisions to the Canadian National Accounts" Statistics Canada Working Paper STC4370E.

CHALLENGES ON MEASURING MIGRATION IN ALBANIA

MAJLINDA NESTURI, INSTITUTE OF STATISTICS
mnesturi@instat.gov.al

Abstract

Emigration was one of the major reasons of population decline in Albania between the 2001 and the 2011 censuses. Given that Albanian emigrants have not systematically registered when leaving the country, and because there have been few incentives to register their departure with the local authorities, the use of 'indirect methods' was the possible way to produce an estimate of the individuals who have emigrated between 2001 and 2011. According to INSTAT's indirect estimations, during this period about 481,000 Albanians left the country. On the return side, Census data revealed that about 139,827 Albanians returned to the country during 2001-2011.
Apart from the 'indirect methods' used for calculation of net-migration, this article puts its focus, as well, in Labour Force survey as a new approach on measuring migration.
Migration remains the main component which drives the population changes. For this reason, its measure is the main challenge encountered when calculating the annual population.

1. INTERNATIONAL MIGRATION

International migration has been a major force in population change in the past two decades. The main reason for the population decline observed in the last two censuses was the large-scale emigration of Albanian citizens. In Albania, as usually elsewhere, migration particularly involves young adults, who migrate for reasons of employment and education opportunities, and for family reasons. Consequently, youth particularly figure in migration statistics and were main contributors to the recorded population decline in Albania.

Emigration was the main reason for the declining population in Albania prior to 2001, and between 2001 and 2011. The abolition of emigration barriers and the social turmoil in the 1990s generated successive waves of emigration in these years. Several hundreds of thousands of Albanians left the country (e.g. Carletto et al. 2006), a loss that was not offset by immigration and natural increase. This massive emigration was the main cause for the complete reversal from a high annual population growth in the 1979-1989 inter-census period of 2.0 percent, to a negative growth (-0.3 percent annually) in the subsequent 1989-2001 inter-census period. In absolute terms, the population increase of almost 600 thousand persons that was observed in the inter-census period before 1989, turned into a decrease of more than 100 thousand in the one after 1989. The 2011 census conclusively showed that emigration was the most important factor in the net population loss of 269 thousand persons between 2001 and 2011, accounting for 8.8 percent of the 2001 population. According to INSTAT indirect estimations, during this period about 482 thousand Albanians left the country, a number that was only partially compensated by immigration and natural growth.

Part of the reason for the continued emigration may be sought in the changing characteristics of Albanian emigration. It is thought that during the 1990s migration was dominated by young single men looking for work abroad, in a time where Albania was in turmoil. In the first decade of this millennium migration seems to have diversified: migration is still dominated by young adults, but in addition to men, women increasingly go abroad as well. The share of children has somewhat increased too, and this is a clear indication of the process of family reunification and marriage migration that often follows the earlier 'pioneering' migration of a first male-dominated wave. Such a process of family reunification and marriage migration tends to enable continued high migration levels at least for some years after the first wave of migration.

1.1 EMIGRATION – THE INDIRECT METHOD

Emigration is one of the major reasons for the declining population in Albania between 2001 and 2011. Given that Albanian residents are not systematically registered when they leave the country, and because there are few incentives to register a departure with the local authorities, there are no reliable measures of the number of individuals who have left Albania. It is possible, however, to use indirect methods to produce an estimate of the individuals who have emigrated between 2001 and 2011. This indirect method consisted on comparing data from both the 2001 and the 2011 Censuses. Over that decade, the population of Albania decreased with 8.8 percent, from 3,1 million to 2,8 million. In order to estimate to what extent external migration was responsible for this decline, the population was projected forward from the 2001 census, using revised figures on deaths and births from the civil register, and assuming no external migration happened (Migration in Albania, 2014).

In a first step, the population as of January 2001 is taken as the basis. Both age and sex of individuals are considered. In a second step, information on live births was taken for each year, taking into consideration the number of males and females born. In a third step, the number of deaths by age and sex are calculated using the specific death rates based on the population projections of the 2001 census (Population Projections for Albania, INSTAT edition, 2004). For each age group, the number of deaths was subtracted from the population in 2001. For the population aged zero, the number of live births in the preceding year was taken into consideration. This procedure was then repeated for each year until January 2012.

The outlined procedure calculates a closed population which does not take into account migration. To calculate the number of emigrants, it is necessary, as a next step, to compare the closed population, as calculated above with the census population. First, the estimated population as of 1 January 2012 (i.e. the closed population) was extrapolated to 1 October 2011 to match the population on the same date (1 October 2011). The difference between these two figures is the estimated net migration.

1.2 RETURN MIGRATION

Given the strong increase in the number of Albanians residing abroad over the decade, return migration has not been very substantial. Nevertheless, in the 2011 census a total of more than 100 thousand people reported that they had returned from residence abroad since 2001. Census analysis demonstrated that returns have been on the rise every year, in particular after 2008 (Migration in Albania).

While some of these return migrants settle in Albania permanently, for many of them the return is temporary in nature. In this sense, the return migration captured in the census is a snap-shot of on-going circular migration. Moreover, the migration figures are based on the internationally accepted definition of an international migrant, as someone who stays abroad for at least one year.

2. DIFFICULTIES MEASURING MIGRATION FLOWS

International migration has become an increasingly important issue for many countries in the world, affecting countries of both origin and destination. Very little is known about the actual number of annual migrants throughout the world, and the scarce information available is contradictory. Today, countries typically rely on their own definitions of what constitutes a "migration". This creates inconsistencies among international data and makes it challenging to understand the process by which people move across national borders.

It has often been pointed out that statistical information on migration is insufficient and incomplete in terms of availability, harmonization of concepts and definitions and accuracy. It is widely recognized that migration data are not easy to collect and that data collections systems, as well as definitions, used to define migration events vary significantly.

Consistent migration flow data are needed for a range of reasons that are critical to the development of a reliable evidence base for migration policies and research. One of the main obstacles is the absence of a standard definition for measuring migration flows. There also needs to be communication and sharing of data between countries of origin and destination, and procedures to assess the reported migration flows.

In Albanian case, measuring migration flows is the main challenge. As introduced above, the only data available concerning migration was the Albanian Population and Housing Censuses, which gives an estimation concerning migration stock within inter-censual period. As it is known using censuses on measuring migration has its own difficulties or limitations. Population and housing census is carried out every 10 years, accommodates only small number of questions and it cannot capture all migration events like entire families that moved as well as seasonal/temporary, circular (night prior to census date). Therefore, it cannot constitute a source of annual statistics on international migration. Moreover, in absence of the administrative data, INSTAT started to focus on social surveys such as QLFS. INSTAT is evaluating also other ways on measuring migration based on the other countries experiences. These methods consist on conducting a specific survey and using mirror statistics. Mirror statistics, sharing of data between countries of origin and destination, will be a good solution if there will be a harmonized definition, as well as defined procedures to assess the reported migration flows.

3. MEASURING MIGRATION USING LABOR FORCE SURVEY

The Quarterly Labour Force Survey is a household based survey. All individuals aged 15 years and over in the selected household are subject of labour force survey. The sample is based in a two-stage sampling procedure. In the first stage are selected the geographical areas with a proportional probability to the size of the enumeration area. In the second stage within each of the geographical areas (once selected in the first stage) are selected a fix number of 8 households by equal probability systematic sampling method. In 2014, for the first time, in the LFS quarterly survey was incorporated a new module for measuring internal, international as well as returned migration. This information through this module is collected on quarterly basis but the data are processed and published on annual basis. The new migration module has been incorporated to LFS questionnaire in 2014; this was a pilot test to check if everything worked. According to LFS 2014, the data collected from this module had some problems, the number of emigrants was underestimated (very low), while on the other hand the returnees was better captured. In 2014, the migration questions had as reference the 2011 census year, in order to capture the entire post-census period, while in 2015 and further on the reference period was the previous year. This module was added to the roster of the questionnaire, this is a part of the questionnaire that collects basic demographic information for all

the household members, and not only for those that are subject of the survey. Data processing consists of several procedures and controls, and then the respective data are used to calculate the population on January 1. The quality of the data collected has continuously been improving and the coverage is more complete on both sides, for emigrants and returnees.

STRUCTURE OF ECONOMIC INDICATORS IN THE PRIVATE NON-FINANCIAL SECTOR YEARS 2010-2016

VALMIRA BEBRI, INSTITUTE OF STATISTICS
vbebri@instat.gov.al

ETUGERT LLAZI, INSTITUTE OF STATISTICS
ellazi@instat.gov.al

Abstract

Various studies show that a country's economic performance is affected by its economic structure. In order to help researchers and policy makers, this paper will focus on studying the economic structure for the private non-financial sector in Albania. The results show that the structure of economic indicators is generally stable despite the annual change of each indicator. The study was carried out based on the results of the annual structural business survey, realized by INSTAT.

KEY WORDS:

Structure of economy; Private non-financial sector; Employees; Value added; Investment; Gender statistics

1. INTRODUCTION

The economic stability of a country is noticed in its economic structure. To understand the economic structure is analysed the behaviour of the sectors that constitute the economy. The study includes all active enterprises, for all legal forms that produce goods and services for the market, with the exception of agriculture, forestry and fishing activities, financial and insurance activities, public administration and defence, compulsory social security, as well as the activities of international organizations.

This article analysed the annual structural business survey data for period 2010-2016, which collects statistical data on basic and macroeconomic indicators, for around 16 thousand enterprises. Basic indicators include data on economic activities, number of enterprises and employees as well as revenues, expenditures and investments. All indicators are expressed in current prices. Macroeconomic indicators are output; intermediate consumption and value added which derive from the basic indicators.

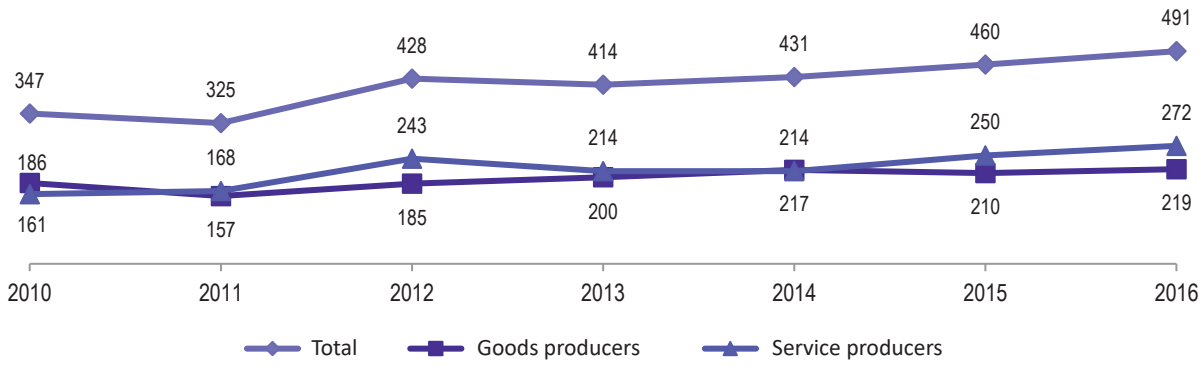
2. INDUSTRIAL OR SERVICE ORIENTED ECONOMY?

The most important macroeconomic indicator is the value added. For the period 2010-2016, for every 100 ALL of value added that enterprises create in the economy, 48 ALL on average is left by the producers of goods and 52 ALL by the services producers.

The analysed data show, as in Figure 1, that in the last three years there is a continuous increase of this indicator, reaching 491 billion ALL in 2016, where services providers have left 272 billion ALL of value added in the economy, while 219 billion ALL are left by the producers of goods. Different situation is presented in 2013 for services providers, where they have contributed negatively to the value added. The impacts to a drop of 3.3% of value added in 2013 have been provided by services providers which contributed by -6.8 percentage points and the contribution of the producers of goods by +3.4 percentage points.

Value added has been the highest decline in 2011 by -6.4%, where the producers of goods contributed by -8.5 percentage points, impacted by bad performance of the electricity sector and the fall of construction activity, against the positive contribution of service producers by +2.1 percentage points.

Figure 1: Value added (billion ALL) by goods and services producers, 2010-2016



To understand better the fluctuations of value added over the years, it was analysed the behaviour of two other indicators; Output and Intermediate Consumption.

Output is an activity carried out under the control, responsibility and management of an institutional unit (enterprise) that uses inputs of labour, capital and raw materials to produce outputs of goods and services.

Referring the data that are analysed, the value of output has been steadily increasing reaching ALL 1,164 billion in 2016, as shown in Figure 2. For the period 2010-2016, this indicator has increased by an average of 6.7%, impacted by service producers that by themselves increased on average with 9.2% and producers of goods that increased by an average of 5%.

For service providers, the negative growth rate of output is observed only in 2013 with -1.1%, by contributing with -0.5 percentage points to the total increase of output by 0.9%.

Intermediate consumption consists of goods and services consumed as inputs by a process of production, excluding fixed assets whose consumption is recorded as consumption of fixed capital. The goods and services are either transformed or used up by the production services.

Intermediate consumption has been on a steady growth, as shown in Figure 3, where in some years the growth rate of this indicator exceeds that of the value of output, by impacted negatively in value added.

A decline in intermediate consumption is noted for the producer of goods for 2013 and 2014, respectively with -0.3% and -2.3%, contributing negatively to the total annual change of this indicator by -0.2 percentage points for 2013 and -1.4 percentage points for 2014.

Figure 2: Value of output (billion ALL) by goods and services producers, 2010-2016

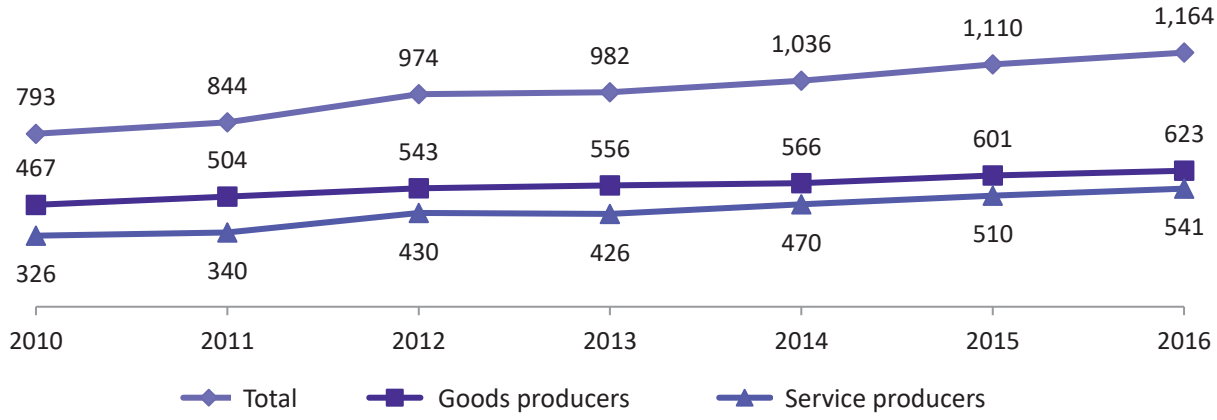
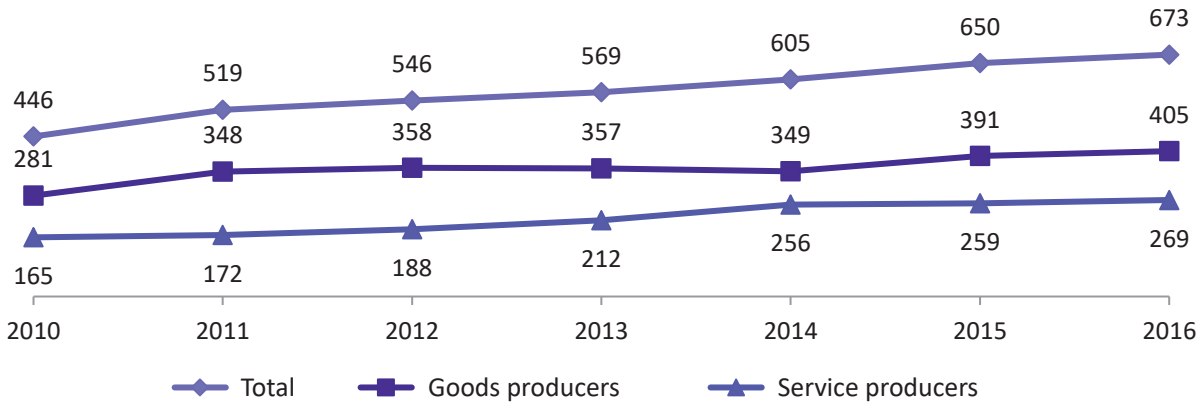


Figure 3: Value of intermediate consumption (billion ALL) by goods and services producers, 2010-2016



3. IT'S EASY TO SERVE...

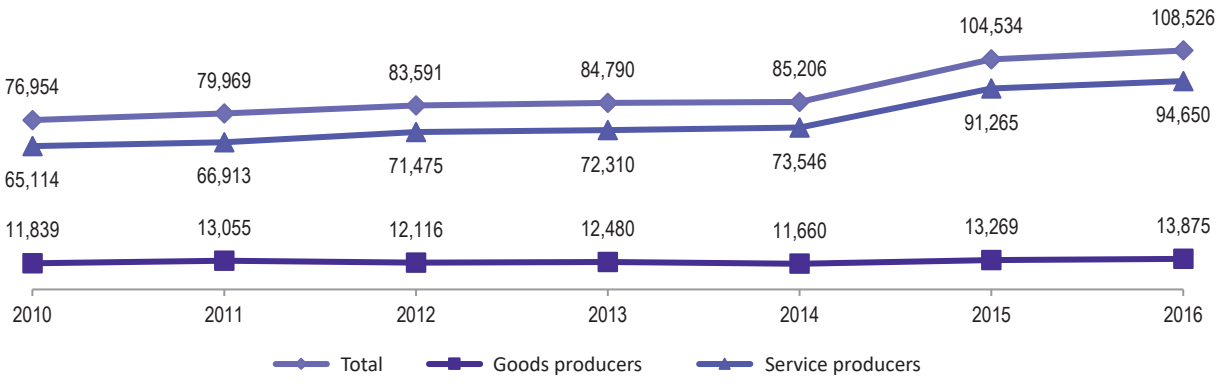
The data analysed on this article show that enterprises operating in our economy are mostly service oriented. The structure of the number of enterprises for the years 2010-2016 shows that the sectors that have expanded more are "other services sector" (20.0% in 2016 from 12.5% occupied in 2010) and "accommodation and food service sector" (18.5 % in 2016 from 16.5% in 2010). While the shrinking sectors in percentage are "trade" (41.5% in 2016 from 44.7% in 2010), "transport and communication" (7.2% in 2016 from 10.9% in 2010), "manufacturing industry" (8.6% in 2016 from 10.4% in 2010) and "construction" (3.1% in 2016 from 4.1% in 2010), as shown in Table 1.

Related with the number of enterprises as an indicator, the year 2015 has the highest increase in percentage of this indicator by 22.7%, influenced also by the government's initiative against informality. Service providers, throughout the years, have had a steady increase in the number of enterprises. The goods production enterprises, decreased their number for 2012 with -7.2% and 2014 with -6.6%, by contributing respectively -1.2 percentage points and -1.0 percentage points to annual changes of this indicator.

Table 1: Structure in number and percentage of enterprises by economic activities, 2010-2016

Economic activity	2010		2011		2012		2013		2014		2015		2016	
	No	%	No	%	No	%	No	%	No	%	No	%	No	%
All market producers	**76,954**	**100**	**79,969**	**100**	**83,591**	**100**	**84,790**	**100**	**85,206**	**100**	**104,534**	**100**	**108,526**	**100**
Goods producers	**11,839**	**15.4**	**13,055**	**16.3**	**12,116**	**14.5**	**12,480**	**14.7**	**11,660**	**13.7**	**13,269**	**12.7**	**13,875**	**12.8**
Mining & quarrying	409	0.5	461	0.6	513	0.6	528	0.6	619	0.7	626	0.6	498	0.5
Manufacturing	8,013	10.4	8,328	10.4	7,731	9.2	7,912	9.3	7,015	8.2	8,137	7.8	9,337	8.6
Electricity, gas, water supply & waste manage.	294	0.4	441	0.6	409	0.5	542	0.6	531	0.6	679	0.6	656	0.6
Construction	3,123	4.1	3,825	4.8	3,462	4.1	3,497	4.1	3,495	4.1	3,827	3.7	3,384	3.1
Services producers	**65,114**	**84.6**	**66,913**	**83.7**	**71,475**	**85.5**	**72,310**	**85.3**	**73,546**	**86.3**	**91,265**	**87.3**	**94,650**	**87.2**
Trade	34,385	44.7	34,724	43.4	36,918	44.2	36,674	43.3	36,357	42.7	45,093	43.1	45,041	41.5
Accommodation & food services	12,687	16.5	12,773	16.0	13,678	16.4	13,784	16.3	14,637	17.2	18,586	17.8	20,043	18.5
Transport, information & communication	8,389	10.9	8,671	10.8	8,255	9.9	8,492	10.0	8,176	9.6	7,365	7.0	7,838	7.2
Other services	9,654	12.5	10,745	13.4	12,625	15.1	13,360	15.8	14,377	16.9	20,221	19.3	21,729	20.0

Figure 4: Number of enterprises by goods and services producers, 2010-2016



4. WHICH ACTIVITY EMPLOYS MORE?

During 2010, goods producers' enterprises employed an average of 10 people while service producers employed 2 people. This indicator on average for 2016 is 12 people employed for producers of goods and 3 employed in service producers. For the whole period of study, the data shows that the highest number of employees for the enterprise is the electricity, water and waste management sectors, while the lowest number is in the trade and accommodation sector and food service sector.

The number of employees has been growing steadily over the years for goods and services producers, despite decreasing the number of enterprises in good producers during 2012 and 2014. The highest increase of the number of employee's indicator is in 2015 (+17.4%), with a corresponding contribution to the producers of goods by +4.2 percentage points and services producers by +13.2 percentage points.

5. FEMALE, FRAGILE LABOUR FORCE

Although the data taken in the study show that there is an increase of employed women in almost all economic activities, there are still gaps with the employment of men. In 2010, only 29.9% of the employed were women while in 2016 the share of women employed was 37.3%, as shown in Table 2. The highest female participation is observed in the manufacturing industry (especially during the last three years of the study, because of the government support for industry of goods for processing abroad), while the lowest percentage in the mining & quarrying industry.

Figure 5: Number of employees by goods and services producers, 2010-2016

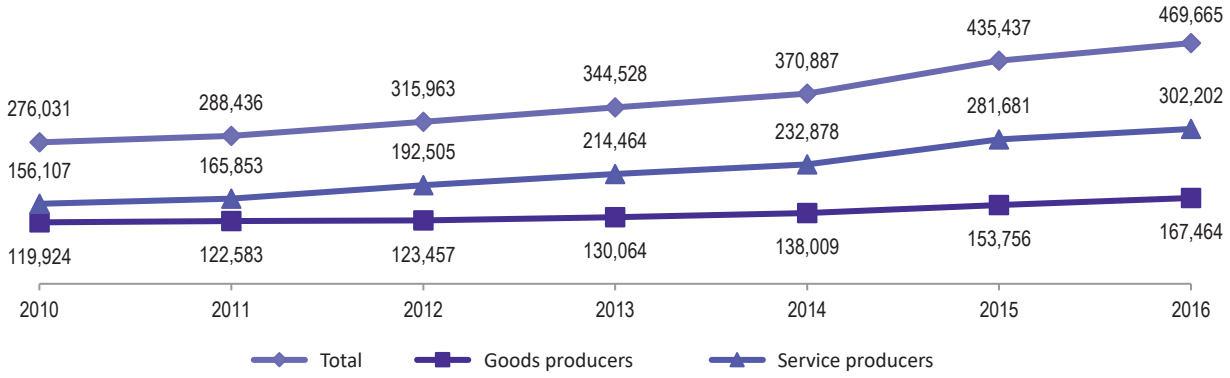


Table 2: Percentage of women employed by economic activity, 2010-2016

Economic activity	2010	2011	2012	2013	2014	2015	2016
All market producers	**29.9%**	**30.6%**	**29.5%**	**28.4%**	**37.3%**	**35.8%**	**37.3%**
Goods producers	**32.6%**	**32.5%**	**31.8%**	**29.5%**	**39.0%**	**37.0%**	**39.7%**
Mining & quarrying	9.6%	8.8%	9.0%	6.8%	10.9%	10.7%	11.9%
Manufacturing	48.6%	51.4%	49.0%	45.4%	58.8%	55.0%	59.0%
Electricity, gas, water supply & waste manag.	31.5%	20.3%	24.2%	17.6%	25.7%	24.5%	18.4%
Construction	9.9%	10.0%	9.9%	10.8%	14.3%	12.7%	12.2%
Services producers	**27.9%**	**29.1%**	**28.1%**	**27.8%**	**36.3%**	**35.2%**	**36.0%**
Trade	29.6%	29.6%	27.4%	31.1%	36.8%	35.3%	36.8%
Accommodation & food services	31.8%	34.2%	31.8%	28.6%	36.5%	30.5%	33.4%
Transport, information & communication	16.0%	18.8%	15.9%	15.8%	25.6%	27.7%	26.6%
Other services	32.1%	32.4%	33.9%	28.5%	40.0%	39.7%	39.7%

6. WHERE ARE THEY INVESTED?

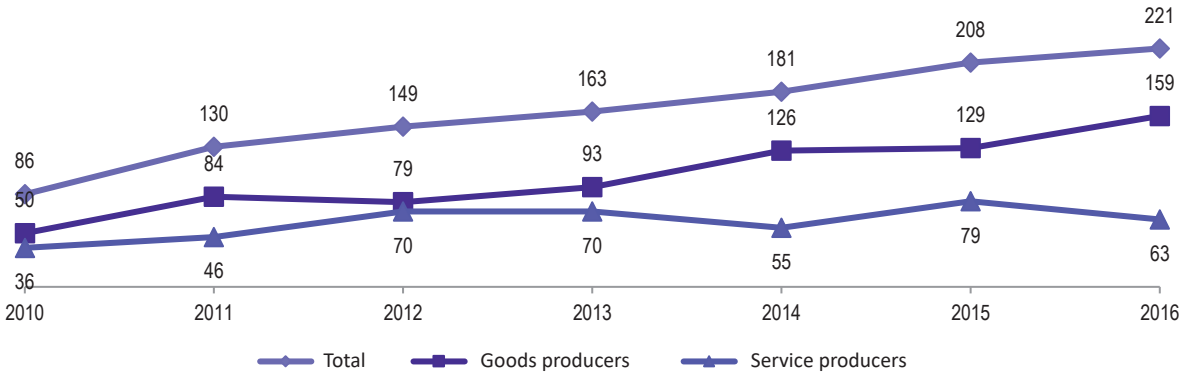
Investments during the reference period include the goods, whether bought from third parties or produced for own use, having a useful life of more than one year including non-produced tangible goods such as land. The value of investments has been increasing in recent years influenced by major infrastructure and energy projects.

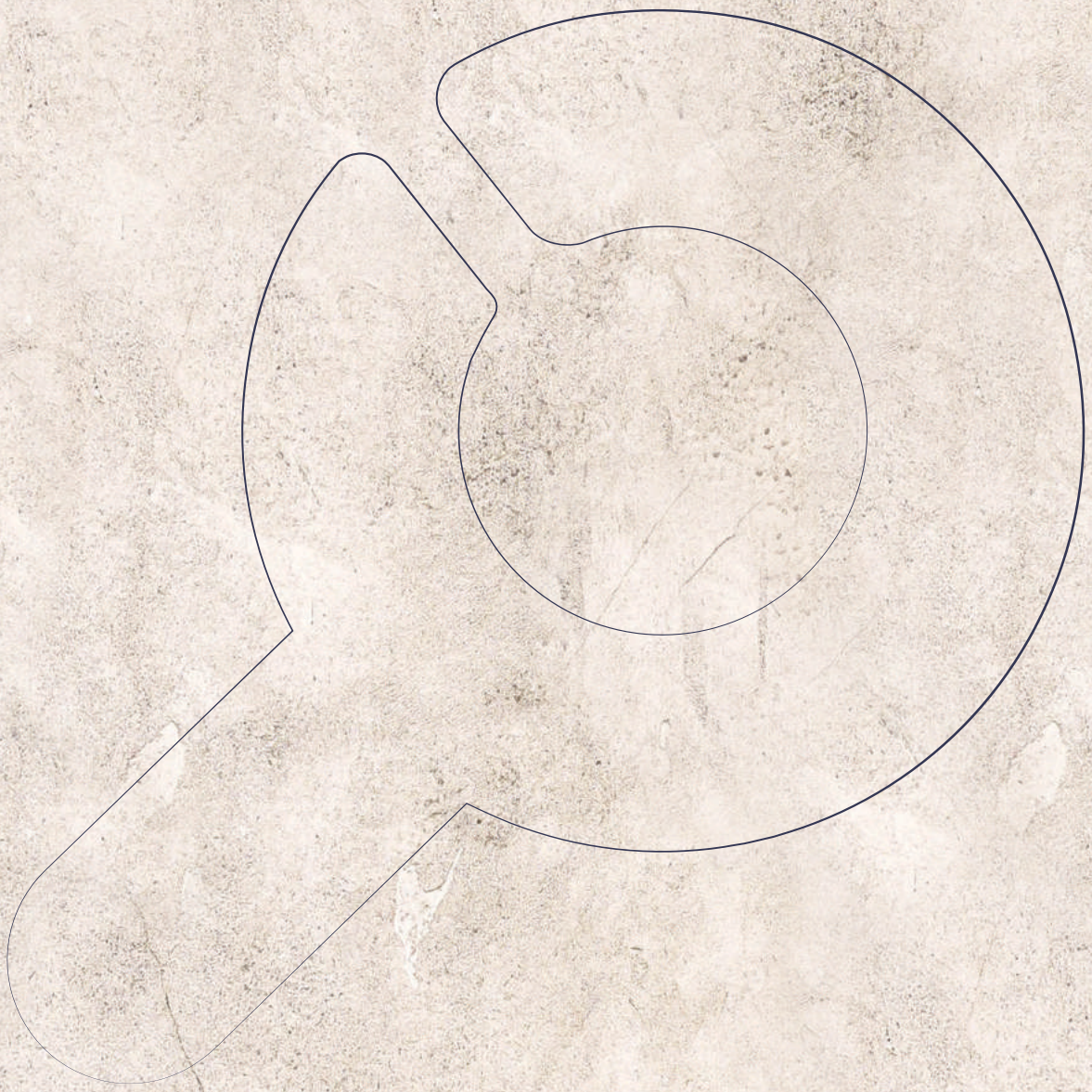
Based on the analysed data for the producers of goods, the investments made by them show a decline in 2012 with -6%, by contributing negatively to the total investments with -3.9 percentage points. The investment done by services producers, has declined for the years 2014 and 2016, by contributing negatively to the annual change of this indicator by -9.4 percentage points and -8.1 percentage points respectively.

The results of analysis the investment structure over the years, according to the types of investments, show that companies invest more in installations, machinery, equipment, buildings rather than in land. In 2016, an enterprise invests approximately 2 million ALL from 1,1 million ALL invested on average in 2010.

As a conclusion of this study we can say that the structure of the Albanian economy has not had substantial changes during the years. From the analysis of the indicators, results show a slight increase of the weight of service producers in the economy compared to the producers of goods.

Figure 6: Value of investment (billion ALL) by goods and services producers, 2010-2016





INSTAT
Instituti i Statistikave