

Register 382 thousand employees were registered. Nor there was significant difference between the LFS and Census data.

One of the main findings in this context is that the HCSO have to rethink to include data on employer, workplace or activities in the case of a future Census questionnaire or this information can be obtained in other ways.

#### **4.2. Matching Population and Agriculture Census data by Addresses and Demographic Variables**

Prior to the 2001 Population Census (PC) a General Agricultural Census (AC) was conducted in April 2000. This is why the management of the HCSO decided to start the Population Census from 1 February 2001 instead of the usual 1 January in order to avoid collision between the two huge operations. The farms (agricultural households) are identified by addresses. The AC form contained a few demographic and employment variables about the persons living in the household dealing with agriculture. These variables are, among others, the following: sex, age, highest degree of education, economic activity, worked days in agriculture. It was reasonable that for deeper analyses the AC data should be combined with more variables coming from the PC. This needed a matching based on addresses and within an address by other variables.

In the HCSO there is neither commercial (e.g. Search Software America) nor in-house developed software in use for address matching. In spite of some efforts to standardize the form of addresses there are various differences in the same addresses used in different data sources, not speaking about the diversity of databases. For example nowadays the change of street names is quite usual.

The task was carried out in the following steps:

i. Creation of clean and common structure of addresses.

It needed manual corrections and adjustments. It is the usual 'standardization' process, the output of which should be a common character string (e.g. when the public area names are to be the same such as Veliko Tarnovo, Main street, 24).

ii. Matching PC and AC records by address.

iii. Matching persons within an address by demographic/social variables.

The matching variables were as follows: sex and age for (almost) exact matching; highest degree of education and economic activity for "loose" matching. The looseness of the matching stems from the fact, that the structure and the length of the code-list of highest degree of education and economic activity are not the same for the two censuses.