Use of administrative data to calculate Labour Costs in NSI of Latvia
(Workshop on Labour Cost Data, Rome, Italy 5-6 May, 2015)

Prepared by Lija Luste, Wage Statistics Section, CSB of Latvia

According to the Law on State Statistics the Central Statistical Bureau (CSB) of Latvia has the right to receive free of charge information from state registers or data bases needed for implementation of the State Statistical Information Program. At the same time the norm allowing the use of administrative individual data on enterprises for statistical purposes is not included in other legislative acts. Despite this, cooperation with administrative data holders lasts for many years based on signed agreements altgough there are still some difficulties preparing new contracts, in particular, requesting data of individuals. There is the Law On Personal Data Protection as well as the Data State Inspectorate that ambiguously treat the use of administrative data for producing statistics. The draft of the new Law On Statistics which was prepared recently, determines the greatest role of NSI in the process of building new administrative data bases, in the further development of those and taking the decisions on termination of maintenance and use. The Law also includes regulations concerning the use and confidentiality of statistical information.

In 2006 the Central Statistical Bureau of Latvia, after long and difficult discussions of several years, finally signed the collaboration agreement with the State Revenue Service on the use of administrative data for statistical purposes. At present the information provided by this institution is the only administrative source for producing wage and labour costs statistics, whereas, data of the State Social Insurance Agency or other institutions are not used.

The main data source both for LCI and LCS is the Quarterly Labour Cost Survey (QLCS) or Survey on Labour carried out by Wage Statistics Section. This survey has a long history. In order to fully meet EU statistical needs, the main changes of the questionnaire were introduced in 2004 when Latvia joined the EU and the first Labour Costs Survey according to the Regulation was carried out. The latest changes in the content of the questionnaire were made in 2008. It should be noted that all elements necessary for the estimation of total labour costs were covered by the survey since 2001. Before LCS 2004 the distribution of employees, their wages and hours worked by full and part working time was not available. The only available distribution was by main and secondary job.

The monthly information on the number of employees, hours actually worked and hours not worked, but paid as well as total wages and salaries in cash is collected by the QLCS questionnaire. The data on the main components of total wages in cash - regular and non-regular wages, payments for annual leave and holidays as well as other labour costs (payments for sick leaves paid by employer, wages in kind, compulsory and voluntary social contributions, imputed social benefits, severance benefit) are compiled quarterly.

QLCS questionnaire is complex and burdensome. It also requests from respondents data on the number of full-time employees dismissed and hired in order to calculate more precisely full time equivalent, basic data for calculation of wages by sex, data on occupied posts and job vacancies

by distribution on main groups of occupations. Furthermore the size of QLCS sample is one of the largest alongside with the sample size of SBS in NSI.

Administrative burden reduction is one of the main goals of the strategy in the whole state as well as in NSI. In order to achieve this goal departments of NSI review all statistical questionnaires every year and discuss the list of available administrative data sources and planned schemes of surveys' samples, devoting particular attention to the reduction of SME burden. Since administrative data of the State Revenue Service have been available for statisticians since 2006, the analysis of data quality and timeliness were immediately started as well as plans for construction of possible estimation methodologies were established. During the financial crisis of 2008/2009, when the budget of NSI was reduced, part of data Collection and Processing Centers employees were dismissed, and dissatisfaction of enterprises caused by the administrative burden intensified this process.

Before 2008 administrative data of the State Revenue Service were used in QLCS only for the evaluation/comparison of variables compiled from surveyed statistical units as well as to compensate errors of non-response for large enterprises. From 2008 until now microenterprises (with the number of employees 6 and less) are not surveyed, but their data are imputed from administrative databases in combination with the established estimation procedure for missing variables. At present the share of such enterprises is almost 80% of all economically active units of the target population covered by QLCS.

The decision to use administrative data for this group of enterprises was made after the analysis of QLCS time series and the results of the Structure of Earnings Survey as well as from comparing them with administrative data. Despite the fact that gross monthly income includes not only regular and non-regular wages but also other income in cash or kind to be taxed (severance and other benefits paid by employer, gifts and presents, stipends and training expenses paid by the employer, various events if it is possible to personify them etc.) the assumption was made that employees of this group of enterprises receive mostly wages in cash and share of other labour costs (except compulsory social contributions) is not significant and administrative data on labour income to be taxed can be used instead of survey data on gross wages and salaries. The data source for the missing variables is the same - QLCS, data of LFS or other surveys are not used. Hours not worked, but paid, regular wages and other labour costs are calculated on enterprise level as shares or rates per employee from the average strata of similar statistical units responded - with the number of employees 7-10 by 2-digit NACE classification. In case no corresponding NACE exists, the average variables of total strata are used. In 2014 the quarterly number of enterprises used for imputation is about 1000, the total number of employees – 8000 (not weighted).

Now the following reports of employers on hired employees to the State Revenue Service are used for the production of statistical variables of microenterprises: the monthly report on statutory state social security contributions, population income tax and entrepreneurship risk state fee; from the end of 2010 – quarterly tax declarations of microenterprises (microenterprise tax payers), too. Both monthly report and quarterly tax declaration include the information of employer identification number (tax payer registration number is common for registration in the State Business register and the State Revenue Service), personal unique ID for each employee

(personal code) and monthly labour income, other available indicators are different. The main and very important difference – there is no information of hours worked in quarterly declaration, i.e., hours are not reported for 40% of employees of this enterprise group. It should be noted that amendment of regulation on declaration had been prepared to make reporting of hours compulsory, but it was rejected because of the obvious burden of micro enterprises.

Regarding reported hours in monthly reports of employers the situation recently has not been so positive. Before 1 July 2013, according to the legislation, hours worked were reported only for employees receiving less than minimum wage. In the beginning of 2013 the share of such jobs was 33% of total. At present hours worked are not being declared only for contractors as well as for some specific groups of employed (senior administrative staff or other persons with autonomous decision-making power, family members).

Another important difference compared to monthly and quarterly administrative data - labour income reported in quarterly declaration is net, i.e., there is income that is paid to an employee ''on hand''. All labour and other taxes are calculated from the micro company tax. Micro enterprise tax payer calculates and pays micro company tax, and does not calculate and pay personal income tax, corporate income tax, risk state fee separately, as well as does not make mandatory contributions. ''Law on micro company tax'' defines the rates of all taxes to be applied according to different criteria. In order to provide gross amounts of wages and salaries as it is defined in the regulations of LCS and LCI as well as in the methodology of national accounts, a method was developed to calculate gross earnings adding only personal income tax, all social contributions calculated from micro enterprise tax are considered as a part paid by the employer. The total personal income tax calculated from micro company tax is distributed as a share of monthly income of each employee from the total income of an enterprise.

Monthly and quarterly administrative data are received 60 days after the end of reference quarter when the first press release on changes of monthly wages and salaries is published. Thus, monthly data of the last month of the quarter and quarterly report of microenterprises for whole reference quarter are not available during the time of data processing. Thereby for imputation data of previous month/quarter are used and updates are made when next quarter's data are produced.

The main steps of administrative data content and quality checking are the following:

1)   overall comparison of the number of data records with those of the previous month;
2)   checking of employees status codes to exclude erroneously included data of physical persons (only legal persons are needed);
3)   summing up gross wages and the number of employees on the level of legal units, adding the summed data to the compiled monthly database of the reference year, comparison/analysis of the summed data. In case of significant changes (no record for an enterprise at all, significant changes of the number of employees) data are linked to the information from the Statistical Business Register and possible reasons of changes are searched (for example, information and date of liquidation);
4)   imputation of missing enterprises' data to be covered in the reference month (previous month's data are used).

After checking and imputation monthly and quarterly data files are merged into one micro file having the same structure, then interior enterprise ID is added as well as other information from the Statistical Business Register (title, NACE, territory code, sector classificatory code – by form of ownership).

Imputation of administrative data and missing variables from average strata is done using a centralized program. For imputation record lines only of those employees are used that have at least 1 euro monthly labour income regardless of the existence of hours worked. Since only the total amount of calculated statutory state social security contributions is available in dataset the employers' part is calculated. The next steps done, but not needed for labour costs data, are the determining of main job, calculating of net earnings (including also corrections of income tax), adding sex from Population Register. Programmer of the Regular Statistics Processing Section calculate missing hours worked for each employee in the prepared micro data file, then impute all variables by summing up on the level of enterprise and calculate other missing variables using data of average strata.

All employees are distributed into two main parts of the questionnaire – having hours worked and do not. Main principle used for the calculation of missing hours worked and imputation is statutory minimum wage, level of labour income and the corresponding hours worked of an individual. Hours for employees with earnings below minimum wage are calculated in a combination of hours worked of full-time employees in the corresponding NACE of average strata and the coefficient calculated as relationship between actual earnings and minimum wage. If NACE does not exist, the average variable of total strata is used. For employees with earnings above minimum wage hours worked of full-time employees of average strata are imputed. For distribution of employees into full or time-parters the threshold of 6,5 hours per day or 32,5 per week in combination with the number of working days in the reference month is applied (the threshold is taken from the data of survey respondents).

As it is mentioned above before July 2013 hours worked were reported only for employees having earnings below the minimum wage. The methodology was updated from the 1st quarter of 2014, taking into account all hours worked available now in administrative data. The main shortcoming of the method applied is the lack of paid hours in administrative data which creates rather inaccurate distribution into full or part time employees needed for LCS as well as for other variables where full-time equivalent is used. Therefore, now it is being planned to use a completely different approach – building up the database of employees or jobs based on all available administrative data containing information of hiring and dismissing dates, lengths of agreement for contractors, lengths of unpaid and child care leaves as well as occupation code of employee.