Overview of employment flash estimation methods

2019 edition

MANUALS AND GUIDELINES

eurostat

Overview of employment flash estimation methods

2019 edition

Manuscript completed in June 2019

Printed by the Publications Office in Luxembourg

Luxembourg: Publications Office of the European Union, 2019

Collection: Manuals and guidelines
Theme: Economy and finance

Print:	ISBN: 978-92-76-08006-0	ISSN 2363-197X	doi:10.2785/239873	KS-GQ-19-005-EN-C
PDF:	ISBN 978-92-76-08008-4	ISSN 2315-0815	doi: 10.2785/044153	KS-GQ-19-005-EN-N

Preface and acknowledgements

Employment is an important variable in national accounts, showing the evolution of the labour market and serving as input when calculating productivity. Employment data form the basis of macroeconomic decisions and are a widely used indicator that reflects how a country's economy is performing. Employment data are also essential for designing and evaluating various government policies.

The ESA 2010 transmission programme allows for a t+60 day transmission deadline. However, many Member States have been granted derogations for timeliness. All derogations will expire on 1 January 2020 at the latest. The quarterly European employment estimates have traditionally been published at t+75 days, after sufficient country data has become available.

The t+75 publication of employment data was behind the timeliness target of 45 days set for the principle European economic indicators (PEEIs)[1] and as such was considered by many data users to be too late. Several Member States have also started to develop employment flash estimates.

In December 2016, Eurostat set up a task force to study the feasibility of producing European employment flash estimates. The results of the task force's work have been published in two statistical working papers.

The first working paper was published in September 2018 and describes the methodology used to produce employment flash estimates for the euro area and the EU. It presents the results of the test estimates using this methodology and the quality criteria applied to evaluate these test estimates.

The second working paper is this document, which presents the methods and techniques used by national data compilers to produce employment flash estimates.

The first European employment flash estimate was published on 14 November 2018, presenting EU and euro area growth rates for the third quarter of 2018. 21 Member States participated by sending country estimates, reaching 91 % of total EU coverage and 95 % of the euro area. The t+75 complete employment estimation has been reduced to t+65 days and is now released together with the data on GDP.

Eurostat wishes to thank all the members of the task force who contributed to the project: Bernadette Boudry (National Bank of Belgium), Jaroslav Zbranek (Czech Statistical Office), Thomas Thorsen/Kris Munch/Sabawon Ghazi (Statistics Denmark), Christoph-Martin Mai (Federal Statistics Office, Germany), Cristina González Fidalgo (Statistics Spain), Franck Arnaud/Bruno Quille (French Statistical Office, France), Nikola Motik/Jasena Torma (Statistics Croatia), Maria Giovanna Piras (Statistics Italy), Jurga Rukšėnaitė/Jaroslav Mečkovski (Statistics Lithuania), Klara Anwar (Statistics Hungary), Johannes Chalupa (Statistics Austria), Peter Čirka (Statistical Office of Slovakia), Yanitsa Petkova (Office for National Statistics, UK), Ani Todorova/Hans Wouters/Jenny Runesson/Véronique Deneuville (Eurostat, National Accounts Production Unit).

Eurostat is grateful for the test estimates provided by the EU Member States not represented in the task force: Bulgaria, Estonia, Luxembourg, the Netherlands, Poland, Portugal, Slovenia and Finland.

This paper has been written by the members of the task force on the employment flash estimates, with the participation of Duncan Elliott (Office for National Statistics, UK). Eurostat particularly acknowledges the work of Johannes Chalupa (Statistics Austria), who organised the drafting.

[1] 2015 EFC status report on information requirements in EMU, available at: http://ec.europa.eu/eurostat/documents/4187653/7065524/EFC-Status-Report-Final.pdf/32a189ce-752c-4545-94e2-d6b114002eff.

Contents

1 Introduction

1.1 Background

Quarterly employment data represent one of the Principle European Economic Indicators (PEEIs). The quality development of these indicators is closely monitored at a policy level and assessed on an annual basis by the Economic and Financial Committee (EFC). Previously, Eurostat's employment estimates for the euro area and the EU in the framework of national accounts only became available at 75 days after the quarter-end. The need for earlier employment data was confirmed by some of Eurostat's key users: DG ECFIN, DG EMPL and the ECB. As acknowledged in the 2015 EFC status report on information requirements in the Economic and Monetary Union (EMU), 'the national accounts based employment indicator still remains well behind the target for timeliness'.

For that reason and based on the fact that some countries already transmit employment flash estimates to Eurostat, in June 2016 Eurostat started a feasibility study by launching a questionnaire in which the Member States were requested to report on the availability of national employment flash data around t+45 days. On the basis of encouraging results of the questionnaire, in the meeting of 14-15 December 2016 the Directors of Macro-economic Statistics (DMES) decided to set up the task force ´employment flash estimates´ whose mandate was to assess whether it would be feasible to produce an employment flash estimate for the euro area and the EU at around 45 days (t+45) or even 30 days (t+30) after the quarter-end. The estimations should mainly be based on the national estimates produced by the Member States.

The task force started its work in January 2017 with focus on these main elements:

- sharing of knowledge between Member States and Eurostat on estimation methods and best practices regarding employment flash estimates;
- preparation of real-time and retrospective national test estimates at/before t+45;
- preparation of real-time and retrospective test estimates for the euro area and the EU at/before t+45;
- development of acceptance criteria to assess the test estimates;
- preparation of a guidance document on recommended methods for flash employment;
- preparation of a methodology for producing euro area and EU employment flash estimates;
- assessment of the results of the feasibility study in an evaluation report and final decisions.

From the technical point of view some conditions were defined:

— regarding timeliness the main target of the feasibility study would be the compilation of employment flash estimates for the euro area and the EU at 45 days after the quarter-end;

— based on the expected reliability and the availability of country data it was decided to limit the exercise to one variable on a high aggregation level: total employment in persons;
— the Member States should transmit to Eurostat unadjusted as well as seasonally adjusted data (either levels or growth rates) so as to enable Eurostat to compile the seasonally adjusted quarter-on-quarter and unadjusted year-on-year growth rates for the euro area and the EU;
— it is crucial that a substantial number of Member States transmit their national estimates to Eurostat and the strict confidentiality of the national estimates both in a test stage and a possible later publication stage is assured.

The results of the task force's work and the employment flash estimates project is documented in two statistical working papers. The first working paper produced by Eurostat focuses on the methodology, results and quality assessment criteria of the employment flash estimates for the euro area and the EU at (t+45) / (t+30) days[2].

The second working paper is this document, which was elaborated mainly by the members of a sub-group within the task force. It focuses on sources, methods and techniques used by national compilers to produce early estimates of national employment.

1.2 Purpose

The purpose of this guidance document is to provide a concise overview of methods and techniques that can be used by national accounts compilers for producing employment flash estimates. Of course, certain topics in the guide may also serve as a good background for both experienced and less experienced data compilers when estimating other macroeconomic indicators or may help users of early employment estimates to interpret their characteristics and quality. The aim is to share the best practices and the technical aspects of estimation methods within the group of people working in the field of quarterly national accounts.

The document structure leads the reader from the description of various sources for employment data and their preliminary analysis, through the estimation procedure ´from source to national accounts´ to the more technical chapter on estimating missing data which was inspired by the similar chapter from the Overview of GDP flash estimation methods[3]. This chapter proposes various forecasting techniques. In the Annexes, the guide presents estimation practices of individual Member States and an overview of useful software and statistical languages.

[2] https://ec.europa.eu/eurostat/documents/3888793/9210927/KS-TC-18-005-EN-N.pdf
[3] https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-TC-16-006-EN-N

2 Sources

2.1 Introduction

In the next sections, several sources used for employment flash estimates will be presented, along with the relevant preliminary analysis procedures. Depending on how a particular country is structured, various data sources may be used for employment flash estimates. Possible sources include:

- labour force survey;
- short-term statistics;
- statistical business surveys;
- administrative data.

The main purpose of continuously collecting and analysing employment data is to provide information on the dynamics of the labour force and monitor employment by activity. The data source used in the estimation procedure depends on the quality, consistency and completeness of the source data.

2.2 Analysis of available data

2.2.1 Labour force survey

Description

Data from business surveys and administrative sources may serve as reliable sources of employment data. However, these sources generally do not include, for example, own-account workers and unpaid family workers. In contrast, the labour force survey (LFS) includes such workers, using a comprehensive labour market survey conducted using households or person samples within certain areas. The aim of the survey is to collect data on the situation of and changes to the labour market and measure the dynamics of both active and inactive individuals, resulting in monthly or quarterly time series. The LFS methodology is consistent with the International Labour Organization's (ILO) and Eurostat's rules and guidelines, which ensure methodological comparability between the EU Member States and thus may serve as a good input source for employment flash estimates. Despite well-defined guidelines and coverage, the LFS is a sample survey and, as such, is affected by sampling errors.

According to the ILO guidelines, individuals are classified into three categories:

- employed (employees and self-employed persons);

- unemployed;
- outside the labour force (inactive), such as school children, many students and pensioners.

Together, the first two categories represent the labour force or the economically active population. The relationship may be expressed as follows:

Population = the labour force + those outside the labour force (inactive)

Labour force = the employed + the unemployed

In accordance with the ILO standards regarding the LFS, employed persons are those who are aged 15 years and more and belong to one of the following two categories:

- individuals who, during the reference week, worked for at least one hour for pay or profit or family gain;
- persons who were not at work during the reference week but had a job or business from which they were temporarily absent.

Unemployed persons are those aged 15 to 74 who were:

- without work during the reference week, i.e. neither had a job nor were at work (for one hour or more) in paid employment or self-employment;
- currently available for work, i.e. were available for paid employment or self-employment before the end of the two weeks following the reference week;
- actively seeking work, i.e. had taken specific steps in the four-week period ending with the reference week to seek paid employment or self-employment or who found a job to start later, i.e. within a period of at most three months.

The LFS defines an employee as an individual who works for a public or private employer and who in return receives compensation in the form of wages, salaries, fees, gratuities, payment by results or payment in kind.

Surveying scheme and sampling design

Although the ILO and Eurostat have published the prescribed methodological guidelines in terms of basic employment definitions[4], every country carries out the LFS on its own, in accordance with the relevant rules, available data sources and technical infrastructure. The principal legal act is Council Regulation 577/1988. Most countries started conducting the LFS annually but at some point in time switched to quarterly levels, covering all weeks of the year and providing quarterly results. The survey usually covers the entire geographical territory of a country and the entire population living in private households, with the exception of collective households.

A key part of the LFS is sampling design, in which persons (households or individuals) are randomly selected with some forms of stratification applied. The sample of persons (households) can be stratified by administrative units (at a given NUTS level) or by size of locality categories in the form of single- or multi-stage systematic sampling from reliable sources, such as central population registers and national address databases, or from a combination of sources. For illustration purposes only, we can take the example of a two-stage random sample in a given country where the primary sampling units can be stratified by a urban/rural division of provinces, while the secondary sampling units can be sampled from selected primary units stratified by municipality size.

Once the random sample has been determined, the interview phase follows, in which a questionnaire is used to collect data. The interview can be conducted on a voluntary basis or may be required by law.

Interviews are carried out using different rotation patterns, such as:

(4) http://www.ilo.org/public/libdoc/ilo/1982/82B09_438_engl.pdf for the ILO and https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_-_methodology for Eurostat.

- each unit is interviewed x times in the course of x quarters. Sampling units are asked to take part in the survey for x consecutive quarters and are then replaced by other units;
- x-(y)-z, in which respondents are first interviewed in x consecutive quarters, then temporarily removed for the next y quarters, before being entered again for the following z quarters, after which they are completely removed from the survey. For example, many countries use the pattern 2-(2)-2, in which respondents are interviewed in 2 consecutive quarters, then temporarily removed for the next 2 quarters, before being entered again for the following 2 quarters, after which they are completely removed from the survey.

As regards technical issues, interviews can be conducted using different survey methods popular in LFS data collection:

- computer-assisted telephone interviewing (CATI), where the interviewer reads questions to the respondents by telephone but answers are recorded on the computer;
- computer-assisted personal interviewing (CAPI), where the interviewer uses his or her computer to read the questions and input the responses;
- computer-assisted web interviewing (CAWI), which is an internet surveying technique in which respondents follow a questionnaire provided on a website and enter their responses into the application themselves;
- paper and pencil interviewing (PAPI), where the interviewer conducts a face-to-face conversation with the respondents, reading the questions and marking the answers on the paper.

Methods 1-3 belong to the so-called computer-assisted survey information collection (CASIC) mode, while method 4 belongs to the paper and pencil mode. Many countries use different methods for collecting data, e.g. data collection by face-to-face interviews is used for the first visit while subsequent interviews are carried out using CAPI and CATI. The survey method used depends on statistical offices' resources and structure and on the technical infrastructure of certain geographical regions within countries.

2.2.2 Short-term statistics

Short-term statistics (STS) describe the most recent developments in European economies and cover four major economic domains: industry, construction, trade and services, using the NACE Rev.2 classification (Statistical Classification of Economic Activities in the European Community).

STS indicators are published as monthly or quarterly indices which show the changes of the indicator compared to a fixed reference year. Employment data are generally speaking available as of 2000 and are transmitted to Eurostat within the 60 days deadline for employment indicators (+15 days for small and medium-sized countries). The employment indicator is provided either as an index or as absolute figures.

STS data are often derived from business surveys. The statistical sources used to establish STS labour input data vary. In some cases special business surveys are used to measure employment, while in others data are collected from administrative sources or a mix of sources. It is difficult to obtain information on self-employed persons from business surveys. This information can be estimated with the aid of LFS data.

The number of persons employed is defined as the total number of persons who work in the unit in question (including working proprietors, partners working regularly in the unit and unpaid family workers), as well as persons who work outside the unit who belong to it and are paid by it (e.g. sales representatives, delivery personnel, repair and maintenance teams).

It includes persons absent for a short period (e.g. those on sick leave, paid leave or special leave) as well as those on strike, but not those absent for an indefinite period.

It also includes part-time workers who are regarded as such under the laws of the country concerned

and who are on the payroll, as well as seasonal workers, apprentices and home workers on the payroll. The number of persons employed **excludes** workers supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the unit in question on behalf of other enterprises, as well as those on compulsory military service.

Unpaid family workers are persons who live with the proprietor of the unit and regularly work for the unit, but who do not have a contract of service and do not receive a fixed sum for the work they perform. This is limited to those persons who are not included on the payroll of another unit as their principal occupation.

2.2.3 Business statistical surveys

This group of data sources covers statistical surveys of businesses conducted by national authorities (e.g. statistical institutes or national ministries) under the relevant national law. The reference period may be short-term (monthly or quarterly) or long-term (annual and lower periodicity). For the purpose of flash estimates the most relevant are short-term surveys. They can also be a major source of data for STS statistics, while annual surveys may be a source of data for structural business statistics (SBS).

Statistical surveys are usually distinguished according to the target group of units examined, i.e. institutional sectors – non-financial and financial enterprises, general government institutions, non-profit institutions serving households (NPISHs), self-employed persons and employers from the households sector. They can cover the units exhaustively or may only survey a sample of units (in which case an appropriate method for grossing up to the whole population is needed). The target units provide the data by filling in the questionnaire (either on paper or electronically) and sending them to the competent institution by a fixed time after the reference period.

The scope of the indicators examined depends on the national legislation and on the budgets of the implementing institutions. Generally speaking, the variables are chosen to meet the needs of the national authorities in the field of state statistics and the European statistical system.

For the purpose of national accounts, the most important indicators are those relating to the production and expenditure of the responding companies. In addition to production indicators, the number of employees can be examined, together with the hours worked and the components of compensation of employees. Data on employees are usually based on the jobs concept, rather than the persons concept and therefore certain adjustments may be needed. Furthermore, companies may not count employed persons who are temporarily not at work but have a formal job attachment (e.g. persons on maternity or parental leave). More information can be obtained by interlinking the survey data with statistical and business registers (i.e. administrative sources).

The advantage of business surveys is the link between production indicators and employment data, resulting in better harmonisation of national accounts data when the surveys are the main source.

On the other hand, the main disadvantage is lower harmonisation at the European level as the surveys are managed under individual countries' legislation and therefore the definitions of employed person may differ. Alternatively, the surveys may not be conducted at the national level at all.

2.2.4 Administrative data

Administrative data are often microdata collected by government institutions. These microdata are usually stored in several registers and are often anonymised when used for analytical purposes.

The registers are unique and quite reliable, and usually contain information about most of the population on an individual level, for example a person's employment status, age, gender and income. The data are usually collected and saved throughout several years.

The available social security data may vary as the scope of social insurance schemes varies from country to country, and from scheme to scheme within the same country.

As described in the 2008 System of National Accounts, social security schemes are social insurance schemes that cover the community as a whole or large sections of the community and are imposed and controlled by governments. The schemes cover a wide variety of programmes, providing benefits in cash or in kind for old age, invalidity or death, survivors, sickness and maternity, work injury, unemployment, family allowances, healthcare, etc.

These schemes generally involve compulsory contributions by employees or employers or both, and the terms on which benefits are paid to recipients are determined by the government. Contributions can also be paid by, or on behalf of, self-employed or unemployed persons.

In this context, social security data can provide detailed information about employed persons with relatively good coverage of the population. These data may be available on a regular basis (monthly or quarterly), a few weeks after the end of the reference period.

3 Estimation procedures: from source to national accounts

3.1 Introduction

Employment data in national accounts may be based on several data sources and Member States use these sources differently. Some countries give priority to labour supply sources, such as the LFS, while other countries focus more on the input from business surveys or administrative data. However, all sources need to be adjusted to fit the concepts of national accounts. This is also an important consideration to take into account when producing a flash estimate, since there is limited availability of sources at an earlier point in time compared to the regular estimate.

3.2 Labour force statistics

According to the current European System of National and Regional Accounts in the European Union (ESA 2010), employment covers all persons engaged in productive activity that falls within the production boundary of the national accounts (section 11.11). Therefore, the estimation of employment should be made by taking into account the residency status of the production unit, since only resident producers contribute to gross domestic product.

The labour force survey (LFS) follows the ILO recommendations. Despite the fact that the ILO concepts share a common framework with national accounts (NA) concepts, both have their own aims and measurement approaches, which may lead to some differences in the final results. For example, as a survey addressed to the population, the LFS estimates the number of persons employed, based on residency within the country's borders, i.e. a national concept, while the concept used in the NA is domestic employment linked to the location of the production unit.

In order to be able to make the transition from the concepts generally used in the LFS to the NA concepts in section 11.19 of the ESA 2010, some items are specified. These and other differences between employment in the labour force statistics and national accounts are detailed below:

(i) Conscripted forces are not included in the LFS, but should be included in the NA under general government services;

(ii) The LFS is addressed to resident households and therefore does not include cross-border workers living abroad, i.e. the group of people resident in other countries who cross the border daily to work in the country. Nor does it include local employees of general government bodies abroad or non-resident crew members operated by resident units. These groups should be included in the NA;

(iii) The NA take into account people who work in the country and participate in producing GDP, although they have not been established in the country long enough to be considered residents. This includes, for example, immigrants and seasonal workers, defined as people resident in other countries who come to work in the country for several months according to the season, but for less than a year;

(iv) In the same way, the LFS collects people residing in the country who cross the border daily to work in another country, seasonal workers or people working for less than a year in another country, resident members of crews operated by non-resident units, local employees of foreign government agencies located in the geographic territory of the country, personnel of the EU institutions or international civilian organisations and the military working in international military organisations located in the country. These groups should be withdrawn from the NA;

(v) Usually, the LFS does not include workers living permanently in an institution and in that case the NA must estimate them;

(vi) The LFS also considers age limits to work, according to legislation, completion of compulsory schooling, etc. Resident workers under the age taken into account in labour force statistics must be included in the NA;

(vii) Although the global concept of persons employed is compatible with both the LFS and the ESA 2010, there may be differences in the distinction between the concepts of self-employed persons and employees[5]. The ESA 2010 defines self-employed persons as the sole owners, or joint owners, of the unincorporated enterprises in which they work, excluding those unincorporated enterprises classified as quasi-corporations. Therefore, owners of incorporated companies, including cooperatives, working in their own companies could be considered employees in the NA, whereas the LFS considers these workers to be self-employed;

(viii) There are also differences in the distribution of the characteristics of employment in the different industries. For example, according to the ESA 2010 criteria, workers provided by temporary employment agencies to other companies must be included in the industry that employs them (division 78 of the NACE 2009), while in the LFS they are classified according to the sector of the establishment for which they are working. However, the distribution by economic activity can also differ because of the different approaches that may be applied in the LFS (usually coding respondent's answers) and the NA;

(ix) NA employment must include employment in the non-observed economy, including illegal activities, where the parties are willing partners in an economic transaction. It is very likely that these kinds of activities are not easily identified in the LFS.

The adjustments mentioned can be summarised in the following equations:

National concept of employment = LFS employment + (i) + (v) + (vi) + (ix)

Regarding the domestic concept, the equation is:

Domestic concept of employment = national concept of employment + (ii) + (iii) – (iv)

Therefore, the results of the LFS have to be complemented by other existing sources in the country. In addition to the different sources of employment already mentioned throughout this guide, some other examples of possible available sources are: (1) population censuses, which provide information about workers living in institutions; (2) structural business statistics, which can give very useful information about the level of employment by industry; and (3) business registers, which provide useful information about the legal status of enterprises by industry. In addition, information about cross-border and seasonal workers may be found in registers managed by different ministries or in surveys relating to this topic. It is also possible to obtain information about the location of the work establishment from the LFS itself.

In some cases, the necessary information to complement the LFS, with the objective of estimating employment levels according to NA concepts, is not available quarterly but on an annual or even multi-annual basis. However, in order to obtain a good estimate of the evolution of employment on a quarterly basis, the LFS data, plus those adjustments that can be made quarterly, may be a good indicator as long as the missing items are of negligible significance and do not have unexpected

[5] A discussion on the proper recording of self-employed persons in the NA is ongoing in the framework of the Quarterly Sector Accounts Task Force and in the National Accounts Working Group of the European Statistical System.

variations.

Even when a country decides that the LFS is the main source of employment estimates, the NA are a synthetic statistic and should be compiled by comparing and combining all the relevant data sources available in the country. This makes it possible to take the best from each source, increasing coherence and obtaining a more comprehensive result.

Moreover, a certain consistency between the estimates of production and employment must be ensured. The necessary balance between the estimates provided by the national accounts should include an adequate link between activity and employment, so as to achieve greater robustness in accounting estimates and improved consistency with other key national accounts variables such as salaries and output. The concept of apparent productivity of the labour factor (value added by occupation) is essential to ensuring this link.

3.3 Short-term statistics

STS employment indicators are often derived from business surveys and do not cover the whole economy. The data need to be completed from other data sources for missing NACE sectors that are not covered by STS. STS employment data are more reliable for sectors where there are few irregular jobs, second jobs or non-market activities, such as industry (NACE B_E), and for selected service sectors.

STS data are usually available with a delay which is too long to enable them to be used directly in the flash estimate procedure (t+45). For this purpose, the indicators need to be updated, for example by using forecasting procedures based on autoregressive distributed lag ADL models (see Chapter 4).

For more information on the adjustments necessary to transform data from business surveys into national accounts concepts, see section 3.4 below.

3.4 Business statistical surveys

Some countries choose to make use of data from business statistical surveys when estimating NA employment. This source covers statistical surveys in enterprises and institutions conducted at the national level that may be focused solely on labour input indicators (number of employees, wages and salaries, hours worked, etc.) or the labour input data may be part of a more extensive dataset beyond indicators of production and business expenditure.

The transition from business surveys to the national accounts concept is very country specific. The surveys differ in frequency, coverage of institutional sectors and industries, types of workers (full-time, part-time) and reporting concept (persons, jobs, full time equivalent). In most countries the business surveys are used as a complementary source to administrative and LFS data or they are used as input data (explanatory variables) for the models in forecasting procedures (see Chapter 4).

Generally speaking, we can say that the following issues need to be considered when adjusting business survey data to the national accounts concept:

- coverage of all domestic institutional units;
- coverage of all economic activities (NACE);
- transition from a jobs concept to a persons concept;
- inclusion of specific types of workers (e.g. armed forces, conscripts, etc.);
- inclusion of the non-observed economy – underground activities.

The coverage of the institutional units surveyed is usually satisfactory for bigger enterprises (above a

certain number of employees) and public sector institutions, while small companies, non-profit institutions and self-employed entrepreneurs are often only covered by a sample survey or are not surveyed at all. Thus, business statistics must be combined with other administrative sources in many cases.

The situation is similar when it comes to the coverage of economic activities. Not all NACE sections are necessarily included in business surveys. Typically, industry, construction, trade and services are well covered when business statistics are used as a source for STS. For activities that are not surveyed, other sources must be used (e.g. farm structure surveys, social security, government treasury, tax returns, the LFS). LFS data cover all activities but generally speaking are less reliable than business statistics because of the higher volatility in the more detailed level of NACE breakdown.

Employment data from business statistics are already in the domestic concept (residents and non-residents working for resident units) but in terms of jobs rather than persons. It is therefore necessary to transition from the jobs concept to persons. This seems to be a difficult issue for countries with a higher number of workers with secondary jobs or seasonal and part-time workers. LFS figures use the persons concept and in addition to main jobs also survey secondary jobs. The sum of employees with main and secondary jobs could therefore serve as a proxy for the total number of jobs. As such, if the business statistics do not include direct information on secondary or part-time jobs, the jobs/persons ratio from the LFS may be used for the conversion.

The ESA 2010 methodology specifies some categories of employees that also need to be included in the national accounts, e.g. armed forces, conscripts, ministers of religion, working students, outworkers and employees on maternity leave. If these categories of workers are not the subject of business surveys, they need to be quantified from other sources, either administrative or LFS data. The figures can be also taken from general government statistics (e.g. the Ministry of Defence or the Ministry of Labour).

Underground or non-registered activities also have to be included in the national accounts. Since business statistics data mainly cover registered employment, some adjustments for exhaustiveness may be needed. Some countries have specialised surveys on underground activities, which can serve as a source for that purpose. LFS data are likely to record a proportion of non-registered activities. Another way is therefore to compare the LFS data, adjusted to the domestic concept (supply of labour force), with business statistics data and administrative sources (demand for labour force). This balancing of the labour force can lead to an estimate of non-registered employment.

Concerning the use of business surveys as a source for compiling flash employment estimates, the main disadvantage is the timeliness of the results for the reference quarter at t+45. If any business statistics data are available at t+45 they can be a valuable source for compiling the flash estimates. However, unfortunately this is not the case in most countries. If only partial information is available (two months of the quarter), forecasting procedures may be used for the missing month. In combination with other sources, business statistics data can show the long-term trends, which may help when estimating flash employment.

3.5 Administrative data

Administrative data enables us to collect information regarding an individual. The information available includes working time, hours paid, numbers of employed persons and employee salaries in a given period of time. The data can be collected monthly, quarterly and/or annually. An example is the reported wages paid per individual employee to tax authorities by the employer. The employer should report retained income tax. The number of employees is therefore quite precise and the working time/hours paid can be derived. From this information it is possible to calculate average employment for a given month, quarter or year. Self-employed persons are sometimes not included in these records and should be estimated from other sources.

The administrative data are collected on individual level and therefore the data needs to be adapted

and adjusted to achieve coherence of the concepts and definitions used in the national accounts system. For example in administrative data we can only analyse and calculate the observed economy by linking different administrative registers together, but we cannot analyse the unobserved economy, which needs to be taken into account for national accounts.

There are several advantages of using administrative data. First and foremost the administrative data is very detailed and the quality is very high. Administrative data is saved in different registers and therefore it is possible to collect a large amount of data with a relatively limited effort. It is both economical and time saving to collect the data and update the existing registers. Administrative data have a consistent method for collecting the data and different institutions are using the same method to collect the same information every year. Each person in the administrative registers has a unique identifying number which is anonymised. Different registers collect different information but the registers can be combined with each other and therefore it is possible to obtain information about individuals from the registers. The same persons are followed over time and one can therefore use administrative data as longitudinal data to follow the individuals across a life span.

The disadvantage of administrative data is that the data is not collected to answer questions of research or evaluation. Administrative data cannot describe attitudes or preferences of an individual as tailor-made surveys can. The data can be used to describe effects, but not why an effort works. Administrative data contains information about individuals on a micro level, which is permanent and can be linked together. This may be unacceptable to some with concerns for potential data leaks or data hacks.

Social security

Information gathered by social security bodies enables the number of persons working to be calculated directly at the company level.

In the context of flash estimates of employment, the dataset available only a few weeks after the end of the reference quarter may be incomplete. In order to complete this dataset, the data for the missing units (employers) could be derived from the data available for the same quarter of the previous year in order to produce a complete dataset. It is then possible to derive year-on-year growth, which can be applied to the non-adjusted series for the number of employees. Another way to use social security data would be to derive the year-on-year growth from a constant sample of employers.

In order to obtain a figure for employment coherent with the national accounts, a distinct estimation has to be provided, firstly for workers not covered by any social security bodies and secondly for undeclared work, in order to obtain an estimate that corresponds to national accounts concepts.

For the number of self-employed persons, the methods above may be not appropriate, owing to the delay in collecting data or to missing data. In this case, the growth rate that could be derived from the available data could be adjusted, taking into account the revisions observed for the previous quarters with more exhaustive data.

The employment figures derived from social security data generally reflect the situation at the end of the quarter or the month. This situation has to be converted into a quarterly average.

Social contribution declarations

National social contribution declarations are different in each country. The following example illustrates the situation in Hungary.

Every employer must produce a declaration of the taxes, contributions and other relevant data relating to the disbursements and benefits for individuals on a monthly basis, in order to account for the advance payments of social security services (SSC), social contribution taxes (SCT) and personal income tax.

1. Figure – Structure of social contributions (example: Hungary)

The declaration has two levels:

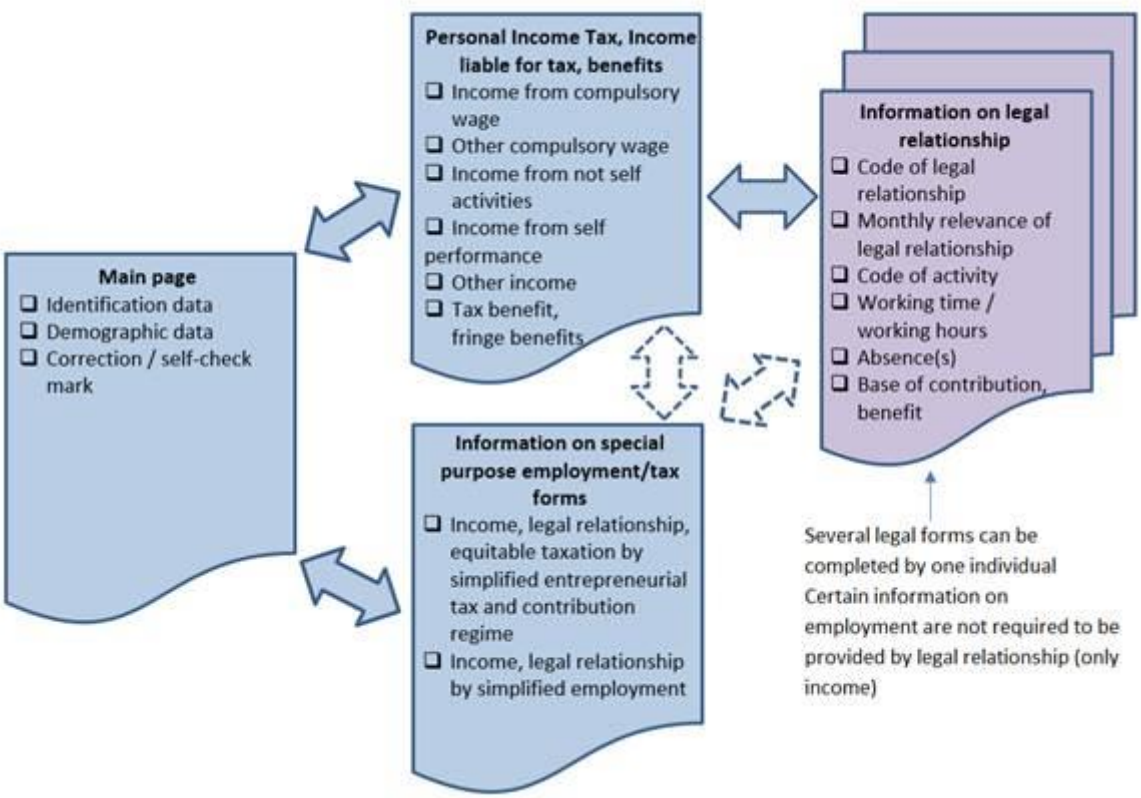
1. The first level consists of two summaries. Firstly, **a summary of the information pertaining to the organisation as a whole**, concerning the SCT, training levy and other employer taxes that require the payment of advance taxes on a monthly basis. Secondly, a summary of the information pertaining to the taxes and contributions deducted from individuals.
2. On the second level, data are to be provided **per individual** (in the case of certain elements: per legal relationship) concerning the taxes and contributions of the disbursements and benefits for the individual. Beyond disbursements, important information is also to be provided regarding the work itself. From the point of view of statistical use, the following should be emphasised:

- the nature of the employment/legal relationship;
- beginning and end of the legal relationship/in the declaration period;
- occupation of the employee/NACE code;
- working time specified in the contract/working hours;
- the nature and duration of absences/SSC-related absences.

In addition, there are also some demographic criteria, such as the employee's nationality, gender, and date of birth.

Moreover, the characteristics (main activity, legal status, total area of the company seat, etc.) of the employer organisation (i.e. the one submitting the tax declaration) that are to be found in the business register extend this information base further.

2. Figure - Structure of social contributions by individuals and information on statistical usability (example: Hungary)



4 Forecasting missing data and adjusting for outliers

4.1. Introduction to a model strategy

Different models can be used to forecast employment data as well as gaps in the time series. This chapter presents methods that help to identify employment data in an econometric way[6]. Since the data for the current period is usually missing, the following sections present forecasting models.

Different methods and procedures can be used in the forecasts. It is possible to calculate employment data in a very simple way, which can lead to good results for a short time. However, in the medium term or where economic development is more uncertain it is recommended to use a procedure based on a type of regression or pure forecasting methods. The most commonly used method for time series analysis in official statistics is the auto-regressive integrated moving average (ARIMA) model. Time series analysis evaluates past data and extrapolates into the future. Detailed information on the use of ARIMA models can be found in section 4.2. Although ARIMA models have many advantages (for example, they are flexible and immediately available), these models are inadequate for detecting exceptional effects or sudden reversals at an early stage as the series are supposed to be weakly stationary. This is no longer the case if there are outliers or in periods of rapid change. As such, it would be beneficial to include other series if they are leading indicators of employment or introduce some indicator variables to capture non-stationary effects, in order to better forecast and understand what happens at the current end of the series. Auto-regressive distributed lag (ADL) models and non-deterministic models should also be tested to check if they perform better than other models. Detailed information on these models can be found in section 4.3. At the end of this chapter, the possibility of using structural time series models (section 4.4) and multivariate extensions (section 4.5) is described.

Naive calculation methods can lead to good estimates in the short term. In such models, the estimating technique is kept simple. For estimations in the area of the labour market, the data from the previous period are usually used as forecast without any adjustment. In time series without any seasonality and economic trend, the forecast is equal to the last observed value. The labour market usually has a recurring season (as a result of weather conditions, holidays, fixed recruitment dates or other factors). A suitable naive forecast that takes into account seasonality involves using the previous year's value. However, the series of employment data also usually follows a long-term economic trend. A naive method should therefore use the last observation plus the average change of the last period. A different approach involves assuming that the series of employment data is developing in the same way as another indicator in the same period (for example, GDP).

(6) Methods have already been presented in the document 'Overview of GDP flash estimation methods' available at https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-TC-16-006-EN-N.

In summary, naive calculation methods can be separated into two groups as follows:

- naive extrapolation without related data (e.g. growth rate of a previous quarter);
- direct estimation using naive extrapolation methods (e.g. forecasts obtained using growth rates of related indicators in the same period).

Although naive models are easy to use and often lead to satisfactory results in some single stable economic phases, they should only be used for comparison and provide a benchmark against some more sophisticated models such as ARIMA or ADL models.

For a correct model strategy (use of regressions models, time-series models), simple models (without a related indicator) should be performed first; complex models should be preferred only if they add value. As regards data availability, it often occurs that the indicator has a monthly frequency and only one or two months of the last quarter are available. In such circumstances, quarterly forecasting requires either the additional step of forecasting the missing months or a mixed frequency model. In the first case, it is usually done by employing a pure forecasting method. These methods fall into the class of 'bridge models'. In the absence of related data, pure forecasting methods (e.g. ARIMA models) can be adopted. Times series models also have multivariate extensions in the literature.

Regarding related times series, it is important to remember that forecasting models do not check causal effects. Theoretically, two completely independent data series relating to the content could have a close mathematical relationship, such as the number of employed persons and the number of ants in a garden. A forecasting model does not necessarily need a causal interpretation. However, quality measures like the coefficient of determination show us that related indicators whose content is close to each other usually lead to better results. A simple regression uses one dependent and one independent variable. Models that use a regression to other data series can also be very complex. Some regression models use multiple independent variables. They are based on the assumption that the value of a dependent variable is a mathematical function of the values of these variables. Using the ordinary last squares (OLS[7]) estimator helps in finding the best model. OLS creates the parameters of a linear function of a set of explanatory variables by minimising the sum of the squares of the differences between the observed dependent variable and the predicted variable by a linear function. However, a forecasting model does not automatically get better when as many regression data series as possible are used. To measure the quality of a forecasting model, the econometric analysis offers a variety of tests, e.g. F-test[8] or Wald-test[9]. These tests prove that the use of more indicators in a model does not automatically lead to better forecast results.

Examples of possible explanatory variables or related data on employment series include: the industrial production index (e.g. manufacturing), industrial new orders, business survey results, short-term business statistics, previously published labour market data (e.g. in some countries LFS data, unemployment data), exports, the construction index, tourism indicators and GDP (as in most countries the labour market is a lagging indicator). However, each EU Member State should have some specific indicators that would be important for forecasting trends in employment.

Related indicators/data could be fully available, partially available or totally absent. Where related data is available, the challenge is to find the best way to bridge the gap between available short-term data and employment data. If no related data are available, the problem is reduced to finding the best estimate on the basis of a pure forecasting method. As regards modelling, in the presence of related indicators, the simplest extrapolation of employment can be obtained by using a regression method and by also taking into account possible autocorrelations in residuals and other effects.

OPTIONS AND SUGGESTIONS

When related indicators are available, it is worthwhile testing them in the model. Where only partial

(7) Sen and Srivastava (1997).
(8) Kočenda and Černy (2014).
(9) Wooldridge (2005) or Greene (2012).

information on related data is available, the bridge modelling approach is recommended. In the absence of related data, univariate forecasting methods (for example, ARIMA models, exponential smoothing and structural models) are the best solution.

Multivariate models are possible extensions of the usual univariate models.

A mix of the previous methods can also be used. For instance, it is possible to cast the regression model in a state space framework and then add further complexity to the model, for example by allowing model coefficients to evolve over time or by using a multivariate framework that picks up the dynamics of the series involved more effectively.

In practice, a mixture between the mathematical solutions above and source-based expert calculations are always recommended, especially in exceptional circumstances, e.g. the influence of weather factors or new statutory rules with employment policy-related effects (protection against dismissal, minimum wage, etc.). On the one hand, the time series analysis and the regression approaches provide important support in estimating the prevailing trends and seasonal movements in employment. On the other hand, peculiarities and irregular effects are taken into account in the expert calculations. The two approaches should be performed independently and the two outcomes should be evaluated against the background of general economic performance and merged into a unique final result. This mix of the two approaches ensures a high degree of flexibility and uses potential synergy effects by means of comparisons. Another option is to combine other forecasts, expert estimates and results from other existing statistics, e.g. labour force survey (Chapter 2.2.1 and 3.2), STS (Chapter 2.2.2 and 3.3) or other statistics.

The Eurostat Handbook on Rapid Estimates[10], released in December 2017, provides an overview of the different types of rapid estimates and the statistical and econometric methods for constructing them. It provides details of the methodologies briefly described in this document.

(10) https://ec.europa.eu/eurostat/documents/3859598/8555708/KS-GQ-17-008-EN-N.pdf.

Outlier detection

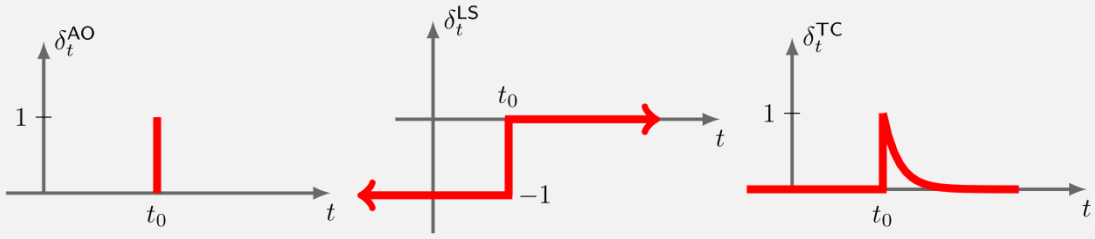
An outlier in a time series is an unusual value that is unexpected given what is believed about the structure of a time series. An outlier might appear due to a one-off event, procedures in an experiment or human factors.

Time series in macroeconomics are often affected by many financial, political and environmental factors (floods, earthquakes and other geohazards). Outliers deserve to be treated in modelling due to their influence on final forecast results. They should therefore either be removed if they relate to experimental or human errors, or treated separately in model specifications – for example, by including them in the model as an explanatory variable. There exist three common types of outlier:

- Additive outliers (AO) occur on a given month or quarter t_0 and affect one observation only. An example for this outlier may be a strike taking place within that period;
- Level shift (LS), which has a shape of a step is an abrupt and permanent change in trend level starting on a given month or quarter t_0. Level shifts occur when there is a structural change in the economy, such as a financial crisis or, for example, a discontinuity due to a change in measurement. One example is an increase in the standard VAT rate during a given period.
- Temporary change (TC) is an abrupt and temporary level change beginning on a given month or quarter t_0 with the level returning to normal at a specified rate. Examples of this outlier are natural disasters (storms, floods, earthquakes) and a strike that significantly hits the economy after which its impact decreases.

Outliers can be identified and treated with regressors in a model adopted for flash estimation. Of course, any that are not significant are excluded from the model specification. Outliers may be viewed in the form of 'functions' and graphics as shown below.

$$\delta_t^{AO} = \begin{cases} 1 & t = t_0 \\ 0 & t \neq t_0 \end{cases} \qquad \delta_t^{LS} = \begin{cases} -1 & t < t_0 \\ 0 & t \geq t_0 \end{cases} \qquad \delta_t^{TC} = \begin{cases} 0 & t < t_0 \\ \alpha^{t-t_0} & t \geq t_0 \end{cases} \quad \alpha \in \langle 0, 1 \rangle$$



The term α in the function δ_t^{TC} denotes a rate of decay back to the previous level.

Ideally, there should be a good explanation for including an outlier, otherwise there may be a danger of over-modelling the series and explaining variation that is not justified, leading to poor forecasts. Identifying outliers at the most recent time points is particularly challenging, as it is difficult to distinguish between the different types of outliers and a general change in the trend.

4.2. ARIMA models

ARIMA stands for auto-regressive integrated moving average. ARIMA models use the past structure of a time series to fit a model that can then be used to extrapolate into the future. It was introduced by George E. P. Box and Gwilym M. Jenkins in 1970. Their approach was that economic time series are non-deterministic processes and ARIMA models are useful for explaining and forecasting.

We consider two simple models for a given time series y_t (quarterly or monthly employment in this case): AR(*p*) and MA(*q*).

The notation AR(*p*) means that the variable of interest is regressed on its own past lagged values abbreviated as auto-regressive model of order *p*. It includes an irregular component ('residual', 'white noise'). Using mathematical language, AR(*p*) can be expressed as

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

$$y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t \quad (1)$$

where φ_i are parameters to be determined, φ_0 is a constant term and ε_t is white noise (variable with mean zero, constant variance and zero covariance).

The notation MA(q) refers to the moving average model of order q, that is

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

$$y_t = \mu + \sum_{i=1}^{q} \theta_i y_{t-i} + \varepsilon_t \quad (2)$$

where θ_i are parameters of a model to be determined as well, μ is a mathematical expectation of y_t and ε_t is, as already stated, white noise.

Both equations (1) and (2) can be combined into an ARMA (p,q) model as

$$y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i y_{t-i} + \varepsilon_t \quad (3)$$

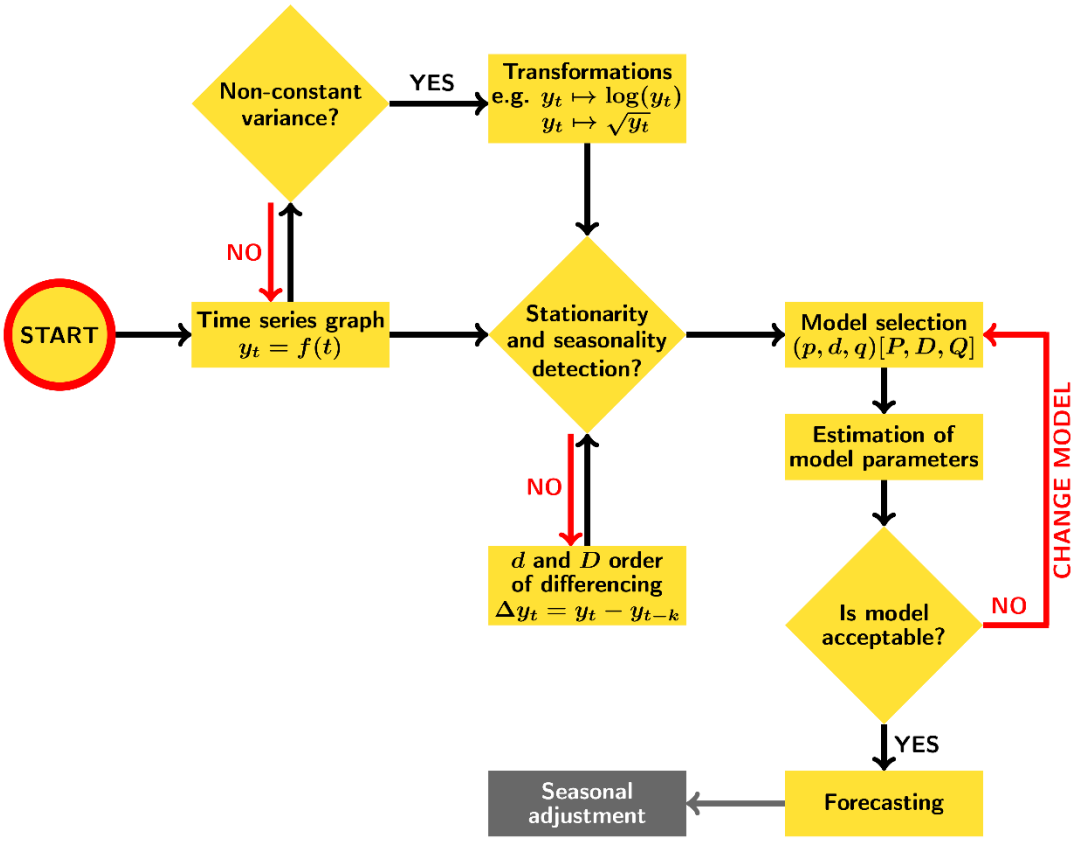
ARIMA plays a filtering role, aimed at separating signal from noise and eventually extrapolating it into the next quarter or month as a forecast value (flash employment in our case). According to equation 3, ARIMA can be viewed as a combination of polynomials depending on the number of auto-regressive terms (p) and number of lagged white noise (q). An ARIMA model includes a polynomial for the order of differencing (d) that may be required to make the time series stationary (note other transformations of y_t may also be required). Most macroeconomic series are non-stationary processes and therefore some transformations have to be applied. This will be described in a so-called Box-Jenkins procedure (see 4.2.1).

The ARIMA (p, d, q) model describes the value of a time series as a function of the order of the auto-regressive (p), integrated (d), and moving average (q) parts of the model, where p, d, and q are non-negative integers determining the order of each polynomial. It may also be important to include seasonal lags in a model to fit a seasonal ARIMA model (p, d, q) (P, D, Q)s, where s denotes the seasonal frequency (12 for monthly and 4 for quarterly series) and P, D, and Q denote the order of seasonal lag polynomials.

Additional regressors can be added to address other causes of non-stationarity, such as outliers. These are sometimes referred to as ARIMAX models or regARIMA models.

4.2.1 The Box-Jenkins procedure

This procedure can be presented in the form of a decision tree, as shown below, with a short overview given for each step:



For the first-order differencing, *k* equals 1, while for seasonal differencing *k* equals 4 or 12, depending on whether quarterly or monthly data are being used.

1. **Time series graph** helps to gain insight into time series trends during certain periods and to recognise seasonality patterns or outliers at the very beginning.
2. **Weak stationarity** is a crucial property for modelling a time series with ARIMA parameters. A time series is weakly stationary if both the covariance and mean are (nearly) constant over time, which means that every shock might occur in time *t* vanishes in some time shift *t* + τ, for $\tau \rightarrow \infty$. In other words, the time series is weakly stationary if the following conditions fulfilled:
    - constant mean over time;
    - constant variance over time;
    - constant covariance over time.
3. **Transformations** are very helpful in making time series stationary. If we take a look at the Box-Jenkins procedure we see that variance stationarity has to be achieved first. Transformation for removing variance non-stationarity can be done using logarithm or square root transformations or any other Box-Cox transformations. Transformation for removing non-stationarity in mean is usually done by taking the first differences and/or by using regressors.
4. **Model selection** is to be carried out once the time series is stationary. At this stage, we choose the best parameters for the ARIMA model (p,d,q) and, if required, seasonal parameters (P,D,Q). Once a model has been specified and fitted it is important to test the

model residuals, which are often assumed to be independent and identically distributed following a normal distribution. This means the model residuals should be approximately normally distributed and show no evidence of autocorrelation or partial autocorrelation. If regressors are included then testing the model residuals for these properties is also important for inference.

In order to choose the transformations and orders of differencing requested to make a series stationary, the major tools used at the identification stage are the autocorrelation function (ACF) and the partial autocorrelation function (PACF). By looking at the ACF and the PACF plots of the differenced series, we can tentatively identify the significant number of AR(p) and/or MA(q) terms.

If the ACF decays slowly and the PACF is represented by a sharp cut-off at a certain lag, we can conclude that the time series follows the AR(p) model. The order of such a model is generally indicated by lags at which the PACF cuts off. More precisely, if the ACF spikes follow an exponentially decreasing envelope and the PACF cuts off after the first lag, we should consider an AR(1) model.

The Akaike information criterion (AIC) is a way of selecting a model from a set of models. The chosen model is the one that has the lowest AIC.

The nature of the data and how it was collected is also a key element. For example, a change in method at a particular point in time requires a level shift or some other type of structural break regressor that should be added manually. If an automatic model suggests a trading day regressor but data concern questions over a week, then a trading day regressor is not appropriate. The structure of a rotating sample might also suggest an order for an ARIMA model.

For users who are not familiar with ARIMA models, some automatic modelling procedures are available in JDemetra+ or in Eviews.

See Annex II for a list of programmes (R, Python) and software.

4.2.2. Estimation and diagnostic checking stage

At the estimation and diagnostic checking stage (assessment of parameter values) we check and diagnose the model, going back to the model identification stage if the previous assumptions are not satisfied. At this stage, the assumptions of the ARIMA model are checked, e.g. the hypothesis of errors being independently and normally distributed, or the statistical significance of coefficients.

We can use maximum likelihood, the method of least squares or Yule-Walker equations to estimate the parameters:

- The least squares estimator of the parameters (ψ, θ, α) is obtained by minimising the sum of squares of model errors.

$$S = \min_{r_i} \sum_{i=1}^{n} {r_i}^2 \ , r_i = \ y_i - f(x, \varphi, \theta, \alpha)$$

- The maximum likelihood method involves finding such estimates of parameters. It maximises the system components and the probability of matching $\mathrm{y_i}$.

$$f\left(y_1, y_2, \dots, y_n \middle| \hat{y}_1, \hat{y}_2, \dots, \hat{y}_n, \sigma^2\right) \rightarrow max$$

- The Yule-Walker equations are the following set of equations

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma^2 \delta_{m,0}$$

where m = 0, ..., p, yielding p + 1 equations. Here, γ_m is the autocovariance function of y_t, σ_ε is the standard deviation of the input noise process, and $\delta_{m,0}$ is the Kronecker delta function.

Regarding transformations and orders of differencing, the major tools used at the identification stage

are the autocorrelation function (ACF), the partial autocorrelation function (PACF), tests on the model residuals, and information criteria such as the Akaike information criterion (AIC).

OPTIONS AND SUGGESTIONS

The ARIMA model selection process should be based on the Box-Jenkins procedure, following all the model selection stages. Beyond statistical significance, evaluation of variables and results should have sound economic interpretability. Moreover, as the selection of ARIMA models is not straightforward, an automatic model selection procedure could be adopted as a first attempt, provided it is accompanied by a reasonable analysis of results and routine revisions of the models selected.

4.3. ADL models

4.3.1. Model selection

DESCRIPTION

Pure forecasting methods (e.g. ARIMA models) analyse past values of the time series. For various reasons, a break in the time series is conceivable, e.g. changes in the business cycle, extreme weather conditions, legal changes in the labour market, etc. Other dynamic models may therefore be preferable. With regard to accurate forecasting, it makes sense to include other variables relating to employment. They should be leading indicators (or earlier published data). One model that takes into account the development of other indicators is the **auto-regressive distributed lag** (ADL) model. ADL is a type of regression model in which lagged values of the dependent variable and current and lagged values of one or more explanatory variables are included as regressors. The inclusion of lagged values is justified by the dynamic nature of most economic processes. A change in the level of an explanatory variable could have an impact on the dependent variables much later than it occurred. For example, if a short-term business indicator is a leading indicator for the development of employment, it may have implications for the growth rates of employment one or more quarters later. Furthermore, changes in the number of employed persons can be revealed through the dependence of the current value on its own past values. The advantage of ADL models is that this information is also used. The development of employment itself is contained in these models.

Although multiple variables can occur with current values and their lags, for the sake of brevity we will present here only a simple ADL (1,1) form, i.e. a model whose regressors include a constant term a_0, the lagged dependent variable Y_{t-1} with regression coefficient a_1 and one explanatory variable X_t taken at lags 0 and 1 and with regression coefficients given respectively by γ_0 and γ_1:

$$Y_t = a_0 + a_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + u_t$$

In the equation above, u_t is the error term with values that have a zero mean and constant variance and are serially uncorrelated (white noise).

Most macroeconomic time series are trended and therefore non-stationary (their mean and/or variance depends on time). The problem with non-stationary data is that the standard regression procedures can easily lead to incorrect conclusions. A spurious regression is a relationship between variables in which two (or more) variables are not related to each other or are wrongly inferred, or in which the model is missing an important variable. It is therefore necessary to test whether the variables are really related. If the variables have a long-run stochastic trend, we say that the variables are co-integrated.

$$Y_t^* = \beta_0 + \beta_1 X_1^*$$

If the variables are co-integrated, it is convenient to express the ADL(1,1) model in an equivalent reparametrised form known as the error-correction model, produced by Engle and Granger:

$$\Delta Y_t = \gamma_0 \Delta X_t - \pi(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + u_t$$

This form has the advantage of including both long-run and short-run information. This is because the long-run equilibrium $Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$ is included in the model, together with the short-run dynamics captured by the differenced term. In this model, γ_0 measures the short-run effect that a change in X_t will have on a change in Y_t. On the other hand, $\pi = 1 - \alpha_1$ is the feedback effect and shows how much of the disequilibrium is corrected, i.e. the extent to which any disequilibrium in the previous period affects any adjustment in Y_t. It provides us with information about the speed of adjustment in cases of disequilibrium. $\beta_1 = (\gamma_0 + \gamma_1) / (1 - \alpha_1)$ is the long-run multiplier. If the variables are in logs then β_1 is the long-run elasticity of Y with respect to X.

The error-correction model is a convenient model measuring the correction of disequilibrium of the previous period, and which also has a good economic interpretation. In the case of co-integration, the disequilibrium error term is a stationary variable. It implies that there is some adjustment process preventing the errors in the long-term relationship from becoming larger. Another important advantage of the error-correction model is the ease with which it can fit into the general to specific approach to econometric modelling, which is in fact a search for the most parsimonious model that best fits the given data sets.

Some explanatory variables to include are:

- the industry production index (e.g. energy, mining, manufacturing);
- tourism indicators (arrivals, overnight stays);
- value added tax collected over quarters – an increase in tax revenues may indicate an increase in sales and thus in employment (assuming constant tax rate over time);
- exports (an increase in exports of goods and services may indicate an increase in employment);
- work construction index (an increase in construction projects requires an increase in work force);
- total loans, as only those who are employed may be eligible;
- quarterly GDP or any of its components.

Not all of the mentioned explanatory indicators will be available quarterly. Some may be published on a weekly or monthly basis. In that case, it is first necessary to aggregate them to the quarterly level.

OPTIONS AND SUGGESTIONS

The final model should be selected after careful comparison of several hypotheses of suitable explanatory variables. It should be possible to give an economic interpretation to the model and all explanatory variables should be statistically significant in terms of t-tests or regression coefficients. If the data are not seasonally adjusted, seasonality should be captured by introducing seasonal dummy variables into the model. The different lags of the variables and different data transformations should be evaluated. When regressors are selected automatically, the significance of regression coefficients should be checked.

4.3.2 Estimation and diagnostics

DESCRIPTION

The Engle and Granger model, known as the error-correction model, can be estimated in four steps:

Step 1: Testing the variables for their order of integration

The first step is to test each variable to determine its order of integration. By definition, co-integration requires the variables to have the same order of integration. Integration of order d means that the

series, after being differentiated d times, becomes stationary. Thus, the first step is to test each variable to determine its order of integration. The Dickey-Fuller and the augmented Dickey-Fuller (ADF) tests are the most frequently used. Other suitable tests include the Phillips-Peron test and the KPSS[11] test. If all variables have the same order of integration, then we can proceed to step two. Where the examined variables follow a different order, non-co-integration is possible.

Step 2: Estimation of the long-term relationship

If the results of step 1 indicate that all variables have the same order of integration, the next step is to estimate the long-term equilibrium relationship and obtain the residuals of this equation. The adequacy of the model should be checked with the help of a test for serial correlation of residuals (e.g. Durbin-Watson test, Breusch-Godfrey test) and a test for heteroscedasticity of residuals (e.g. White test). Moreover, the multicollinearity of the explanatory variables should be assessed. If there is co-integration, the results obtained will be spurious.

Step 3: Checking the order of integration of the residuals

In order to determine if the variables are actually co-integrated, it is necessary to test if the estimated residuals from the long-term relationship are stationary. If so, then the variables are co-integrated. This can be done by performing an ADF test (or another stationarity test) on the residual series. The critical values differ from the standard ADF values, being more negative. If we find that the residuals are stationary then we can reject the hypothesis that the variables are not co-integrated.

Step 4: Estimation of the error-correction model

If the variables are co-integrated, the residuals from the equilibrium regression can be used to estimate the error-correction model and analyse the long-term and short-term effects of the variables, as well as to see the adjustment coefficient, which is the coefficient of the lagged residual terms of the long-term relationship identified in step 2. Finally, we always have to check for the adequacy of the model by performing diagnostic tests for serial correlation and heteroscedasticity of residuals.

A drawback of the Engle-Granger approach is that when there is more than one explanatory variable, there may be more than one co-integrating relationship, and the procedure using residuals from a single relationship cannot accommodate this possibility. This can be resolved by applying the Johansen test, which provides us with the number of co-integrating relationships.

OPTIONS AND SUGGESTIONS

The statistical properties of the model should be verified using a variety of diagnostic tools, covering integration and co-integration of variables, serial correlation, heteroscedasticity and multicollinearity. In addition, a specification test and a test for stability of parameters should be performed, as should tests to establish the order of integration of the variables and their possible co-integration. There is a risk that estimation without diagnostic checking could result in serious misspecification of the model and spurious regression.

4.4 Structural time series models

A popular alternative to ARIMA models is provided by structural time series models. Here, the series is expressed in terms of components of interest such as trends, cycles, seasonal components and the irregular term.

Structural time series models are given a standard statistical treatment within the Kalman filter, which allows for simple forecasting, estimation of missing observations, regression effects, outliers, seasonal adjustment and temporal disaggregation. The main differences between structural time

(11) Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test.

series and ARIMA models are that the latter have the advantage of parsimony in model representation, whereas the former are designed according to the analyst's desire to see given components in the form describing the data evolution according to a given model[12].

OPTIONS AND SUGGESTIONS

The best way to carry out the analysis is to fit a general form (basic structural model or local linear trend) to the time series under analysis, including, when required, a sound strategy for log transformation, outlier detection, treatment of deterministic effects and seasonal adjustment. In the second step, the analysis is addressed towards model selection and diagnostic checking. Finally, when computing the desired forecasts, some elements of automation could also be considered in the procedure for model selection, estimation and forecasting. Another way of simplifying the process is to limit model selection and estimation to the more representative forms.

It is important to avoid fitting a model of deterministic structural components (such as a linear trend or seasonal dummies) to observed time series by means of simple OLS regressions.

4.5 Multivariate extensions

Both ARIMA and structural time series models are models for univariate time series whose use could be extended to multivariate settings. In a multivariate model, both the target and the explanatory variables are treated as a cross-section of time series. Following Harvey (1989, p. 429), under a multivariate setup, it is assumed that the different series are not subject to any cause-and-effect relationship between them. However, they are subject to the same overall environment, such as the prevailing business climate and therefore a multivariate model will seek to link them together. The main advantage of multivariate models is therefore the fact that they overcome the assumption of exogeneity of explanatory variables implicit in regression methods. Furthermore, such models often provide more useful information about the dynamic properties of the series and produce more accurate forecasts. By contrast, their statistical treatment is more complicated as the number of unknown parameters to be estimated increases rapidly with the number of series to be treated together, and their identifiability might become complicated.

The possibilities include those listed below:

- Vector auto-regressive (VAR) models are the multivariate extension of ARIMA models without the MA component. This class of models is popular for its simplicity of statistical treatment. A very good basic introduction to VAR models can be found in Kočenda and Černy (2014), while some advanced topics can be found in Lutkepohl (2005);
- Multivariate extension of structural time series models, i.e. seemingly unrelated time series equation (SUTSE)[13] models. This class includes dynamic factor analysis for which, given a representation of a cross-section of time series into components (trend, cycle, etc.), a specification sharing certain components may exist. In other words, a reduced number of these components is informative for the entire set, simplifying the model specification. When common factors are found in the trend, the model is co-integrated;
- A very interesting way of approaching the problem of flash estimates is to extend the multivariate settings to models handling mixed frequency data. As discussed in previous sections, in the real world data are often available for different time spans (in the example above, data were monthly and quarterly). The possibility of forecasting together in a unique setting could simplify their statistical treatment. In particular, adopting multivariate models with mixed-frequency data overcomes the multi-step procedure of bridge models, and the flash estimate is calculated in one step. Here, missing data at the end of the sample – the

(12) However, many ARIMA models can be written in a state space framework.
(13) Seemingly unrelated time series equations.

ragged edge problem – are estimated together in one step once model parameters have been obtained. For recent contributions see for example the works by Kuzin et al. (2009), Clements and Galvao (2008), Banbura at al. (2013) and Van den Brakel et al. (2016);

- Temporal disaggregation has also been dealt with in the literature in the context of mixed-frequency multivariate time series models. See the contributions by Frale et al. (2010 and 2011) on constructing a euro area monthly indicator of economic activity based on factor models. See also Moauro and Savio (2005) and Moauro (2014) on application to employment based on SUTSE models. In this case, estimates of quarterly data are a by-product as they are obtained through temporal aggregation of monthly estimates;
- This list of methods is, of course, not exhaustive and there are several other possible modelling strategies. These include the nonlinear class of state-dependent models or the switching regime models introduced by Priestley in 1980. Here, the dynamic paths are governed by a set of auto-regressive parameters, a set of moving average parameters and a local intercept, each of them dependent on past information. For a general outline of nonlinear models, the classic references are Priestley (1988), Tong (1990) and Granger and Teräsvirta (1993).

OPTIONS AND SUGGESTIONS

Analysts should select an appropriate multivariate model strategy, according to the data and resources available and the relevance of the exercise. They should then conduct a comparative analysis of the results using more common univariate frameworks, in order to understand the real usefulness or value added of a complex multivariate framework. When a well-established multivariate set-up is used, comparative analysis could be avoided. In any case, a proper preliminary analysis of the available indicators and diagnostic checking are strongly recommended. The use of complex multivariate frameworks without experience and without comparative analysis using univariate methods should be avoided. For other model extensions see Eurostat's 2016 Overview of GDP flash estimation methods.

References

Banbura M., Giannone D., Modugno D. and Reichlin L. (2013), Now-casting and the real-time data flow, Working paper series, No. 1564, July 2013, European Central Bank.

Clements, M.P. and Galvao A.B. (2008), Macroeconomic Forecasting With Mixed-Frequency Data, Journal of Business & Economic Statistics.

Eurostat, European System of Accounts – ESA 2010, Publications Office of the European Union, Luxembourg, 2013.

Eurostat, Labour force survey in the EU, candidate and EFTA countries – main characteristics of national surveys, 2013, Publications Office of the European Union, Luxembourg, 2014.

Eurostat, Labour force survey in the EU, candidate and EFTA countries – main characteristics of national surveys, 2016, Publications Office of the European Union, Luxembourg, 2018.

Eurostat, Handbook on Rapid Estimates, Publications Office of the European Union, Luxembourg, 2017.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2010), Survey Data as Coincident or Leading Indicators, Journal of Forecasting.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2011), EUROMIND: A Monthly Indicator of the Euro Area Economic Conditions, Journal of the Royal Statistical Society - Series A.

Granger, C.W.J. and Teräsvirta, T. (1993) Modelling nonlinear economic relationships, Oxford University Press.

Greene, H. W. (2012), Econometric Analysis, Prentice Hall, Upper Saddle River.

Hussmanns, R., Mehran, F. and Verma, V. (1992), Surveys of economically active population, employment, unemployment and underemployment: An ILO manual on concepts and methods, International Labour Office, Geneva.

Kočenda, E. and Černy, A. (2014), Elements of time series econometrics: An applied approach, Karolinum Press, Charles University, Prague.

Kuzin, V., Marcellino, M. and Shumacher, C. (2009), MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area, Discussion paper series 1 economic studies, N.07/2009, Deutsche Bundesbank.

Lütkepohl, H. (2005), New Introduction to Multiple Time Series Analysis, Springer Verlag, Berlin.

Moauro, F. and Savio, G. (2005), Temporal disaggregation using multivariate structural time series models, The Econometrics Journal, volume 8.

Moauro, F. (2014), Monthly employment indicators of the euro area and larger Member States: real time analysis of indirect estimates, Journal of Forecasting, Volume 33.

Priestley, M. B. (1980), State dependent models: a general approach to nonlinear time series analysis, Journal of Time Series Analysis.

Priestley, M.B. (1988), Nonlinear and nonstationary time series analysis, Academic Press.

Sen, A. and Srivastava, M. (1997), Regression analysis: Theory, methods, and applications, Springer Verlag, New York.

Tong, H. (1990), Nonlinear time series, a dynamical system approach, Oxford University Press.

Van den Brakel, J.A. and Krieg, S. (2016), Small area estimation with state-space common factor models for rotating panels. Journal of the Royal Statistical Society Series A-Statistics in Society.

Annex I Employment flash estimation methods by country

Belgium

1. Sources and their availability/timeliness

Administrative sources (mainly data from various social security institutions) are the main statistical sources used to estimate the number of employees and self-employed persons in Belgium. These administrative data become available no sooner than three months after the end of the reference quarter.

Thus, at 30 days or 45 days after the end of the reference quarter, no administrative data are available.

2. Variables to be estimated

The variables to be estimated are the number of employees and the number of self-employed persons, in order to produce total employment, according to the domestic concept.

3. Description of the estimation process

The estimation process relies on an econometric method to extrapolate the evolution of paid employment using other available information (value added, bankruptcies, paid employment for the previous quarter).

As the number of self-employed persons is quite stable, the assumption is that growth in the number of self-employed persons is similar to the quarter-on-quarter growth observed for the previous quarter.

4. Use of modelling

An auto-regressive distributed lag (ADL) model is used to estimate the evolution of paid employment. The explanatory variables used in this model are as follows:

- number of company bankruptcies (monthly data, 12 months moving average);
- value added in the manufacturing industry by volume;
- total domestic paid employment, quarterly averages.

These variables undergo a log-transformation and are expressed in variation compared to the previous quarter.

The selected lags are:

- quarters Q and Q-1 for the number of bankruptcies;
- quarters Q, Q-1, Q-2 and Q-3 for the value added in the manufacturing industry;
- quarter Q-1 for paid employment.

The output of the model is a quarter-on-quarter calendar and seasonally adjusted growth rate for the considered quarter from which a year-on-year growth rate for the number of employees is derived.

5. Results, publication, revisions

At present, the first estimate for employment for Belgium is published 60 days after the end of the reference quarter using the model presented above.

This estimate is then revised at 120 days after the end of the reference quarter, using administrative data (social security data) for the first time.

Every time annual data are revised, consistent quarterly series are provided in the first release following the yearly publication.

6. Other information

The year-on-year and quarter-on-quarter growth rates obtained for paid employment are compared with the growth rates of value added with a 3-quarter lag in order to assess the quality of the estimate and its coherence with this aggregate.

Czechia

1. Sources and their availability/timeliness

There are three groups of available data sources for the purpose of t+45 employment flash estimates in Czechia:

LABOUR FORCE SAMPLE SURVEYS (LFS)

The LFS concentrates on households living in dwellings chosen at random. The LFS is a continuous survey, whose results are evaluated and released on a quarterly basis.

The LFS data needed for the purpose of estimating employed persons are available in good time. The LFS data are received between 33 and 38 days after the end of the reference quarter. They are the main data source for estimating employed persons (main jobs for employees and self-employed persons). The industry structure of self-employed persons and the number of residents working abroad are also taken from the LFS.

SPECIAL EMPLOYEES DATABASE (SD)

The SD is a special database is maintained by the Processing Methodology and Data Sources Unit of the Czech Statistical Office (CZSO). The database contains information on employees based on information from the Register of the Czech Social Security Administration (CSSA). The data in the database are adjusted according to the results of the relevant statistical surveys (monthly and quarterly surveys of non-financial enterprises) and data from other registers (mainly from the business register). The SD is updated on a monthly basis according to the data received from the CSSA.

The data from the SD are used to estimate the industry structure of employed persons for employees.

ADMINISTRATIVE DATA SOURCES

Administrative data sources, such as data from the Ministry of Labour and Social Affairs and the Ministry of the Interior, are supplementary data sources. These data sources are used to quantify the number of non-residents working for resident units.

2. Variables to be estimated

We estimate the number of persons with a main job, where main job means the job with the greatest number of hours worked (if the person has more than one job). Employment, expressed as the number of employed persons, includes the number of employees and the number of self-employed persons. Employment in terms of the national accounts is estimated in the domestic concept.

3. Description of the estimation process

The estimate of the number of persons is based on the information regarding main job obtained from the LFS, which is adjusted in compliance with the ESA 2010 methodology. These adjustments include conceptual adjustments and exhaustiveness adjustments. In addition to these adjustments, the extrapolation of underestimated or missing data is carried out. Exhaustiveness adjustments represent estimates of the number of persons operating in the non-observed economy.

The transformation of the LFS data from the national concept into the domestic concept is the most important conceptual adjustment carried out. This adjustment is necessary because the data from the LFS correspond to employment in the national concept. Transforming the data from the national into the domestic concept of employment means excluding workers working abroad (in this case Czech workers working abroad for less than one year) and including non-residents working for resident units (foreigners working in Czechia for less than one year).

4. Use of modelling

We model the data on residents working abroad as well as non-residents working for resident units and the non-observed economy for the purpose of estimating the number of employed persons. The number of non-residents working for resident units is estimated using the model of the balance of the number of

foreigners in Czechia. The model uses data from administrative data sources such as data from the Ministry of the Interior, the Ministry of Labour and Social Affairs and the Ministry of Education, Youth and Sports. The results of checks carried out by the Labour Offices with foreign police inspectorates and customs offices in corporations are used as the basis for making expert estimates of the illegal employment of non-residents.

5. Results, publication, revisions

The data on employed persons are published quarterly within news releases together with a preliminary estimate of GDP. The quarter-on-quarter and year-on-year percentage changes in the total employed persons in the reference quarter are published. The quarter-on-quarter data are adjusted for seasonal and calendar effects.

The compilation level of t+45 employment flash estimates represents the data on employed persons compiled for the whole economy, broken down by the 2-digit level of the NACE rev. 2. The employment estimates are estimated separately for employees and self-employed persons. The employment estimates are compiled in the form of non-adjusted data and seasonally adjusted data.

Denmark

The present document describes the method for compiling t+30 and t+45 employments flash estimates at Statistics Denmark. The document describes the background of the project, the method used to obtain data and the results for employment flash estimates t+30 and t+45. Flash estimates are performed using the Danish Working Time Account (WTA) as the basic source and with methods that are as close as possible to sources and methods used for the ordinary calculation of the national accounts. For more information, see the quarterly EU documentation: Documentation for quarterly national accounts[14] and GDP-indicator documentation: GDP-indicator documentation[15].

1. Sources and their availability/timeliness

The main data source to produce the indicator for employment flash t+30 and t+45 is the Working Time Account (WTA), independently produced and published by the division for labour market in Statistics Denmark. WTA is based on a number of administrative registers – primarily Labour Market Account (LMA). These data sources are adapted and adjusted to achieve coherence of the concepts and definitions used in the Danish Working Time Accounts system. WTA is produced in such a way that the concepts are, to the greatest extent within the framework of the register data, adapted to the concepts of the national accounts.

Data coverage in the WTA:
The data in WTA covers employees, self-employed and assisting spouses working in resident Danish enterprises or on resident Danish ships, which is consistent with ESA 2010 (see ESA2010 paragraphs 2.04 to 2.11). In the WTA, developments in employee data from the *Employment Statistics for Employees* are used for projection of levels from the LMA. In the WTA, the latest development in the LMA is used for projecting employment of self-employed and assisting spouses. Data is split in primary job and one or more secondary jobs. The distinction between primary job and any sideline jobs the person may have is carried out using the number of paid hours of work in the jobs. If paid hours are the same, the distinction is made using information on wages. Only data for main/primary jobs are used for flash estimates t+30 and t+45. Employment figures are therefore equal to primary jobs incl. persons temporary absent from work.

Employees

Employee statistics are monthly data on the number of persons with employee jobs. The data are broken down by a person's main jobs and other (sideline) jobs (prioritized according to number of paid hours of work and secondarily according to payment in the job). Only main jobs (incl. persons temporary absent from work) are used in the flash estimates for t+30 and t+45.
For more information, see the documentation for employee statistics: Employment statistics for employees[16].

Self-employed, including assisting spouses

Data for self-employed are "modelled" (projections) based on administrative data. However, no data regarding the reference period are available, so instead projections are used. The development of the last two years data from LMA is used to project the data for reference period for self-employed. The projections are carried out on detailed lines of industry, but at the same time ensuring consistency to a more aggregated level. The projection is merely a continuation of the current trend in LMA, so that employment for self-employed or assisting spouses develops in the same way as between the last two years of LMA data (calculated monthly by means of annual growth rates). The starting point for the calculation of the average employment in the month in the WTA is status information on number of persons employed each day of the month, according to the Labour Market Accounts (LMA). LMA data on self-employed persons and assisting spouses originates from various sources, especially income statistics and business registers. This procedure is a very detailed combination of administrative data.

(14) https://dst.dk/ext/national/Danish-QNA--pdf
(15) https://www.dst.dk/ext/national/GDP-Indicator-Dk
(16) https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/monthly-employment-for-employees

For more information, see the documentation for WTA: Documentation of the Danish Working Time Account[17].

2. Variables to be estimated

In order to calculate t+30 and t+45 employment flash estimates we receive monthly data from WTA. The data used in the NA to calculate the estimates for employment incorporates the same procedures as in the quarterly national accounts. The variables that have to be estimated are the total number of employees and self-employed persons which are based on national accounts concepts. All estimates are based on non-seasonally data and the process is finalized by seasonally adjusting the National Accounts results on a quarterly basis using the same level of detail and the same procedure as in the regular national accounts. This also implies calculation of year-over-year and quarter-over-quarter percentage changes.

3. Description of the estimation process

T+30

The t+30 estimate is calculated based on WTA for two months only and the third month is an estimate (projection) using ARIMA model. Data from WTA is received by 69 industries which are aggregated to 21 industries. The ARIMA model projects the third month by 21 industries for both employees and self-employed, including assisting spouses. The data for two months and the projection for the third month are now aggregated to quarter and the year-over-year and quarter-over-quarter percentage changes can be calculated.

T+45

At t+45 days, an estimate for the total employment for employees for the third month of the quarter is obtained from eIncome. This aggregate is then compared with the aggregated forecast from t+30 and the difference is then distributed across sectors based on sector size as calculated for t+30.

4. Use of modelling

ARIMA modelling is used for t+30 when data for the third month of the quarter are missing. ARIMA modelling is used on non-seasonally adjusted data for 42 industries. The forecast is used together with observations for two months to calculate the average employment for the quarter. The quarterly figures are subsequently seasonally adjusted with the same procedure and level of detail as in the quarterly national accounts.

5. Results, publication, revisions

Statistics Denmark does not yet publish flash estimates, but provides year-on-year and quarter-on-quarter percentage changes to Eurostat for the purpose of calculating the total EU estimate.

The following tables contain results for employment flash estimates for t+30, t+45 and t+2 months, both year-on-year (table 2) and quarter-on-quarter (table 1).

Table 1: Quarter-on-quarter growth rate (%)

	Period			
	2017Q1	2017Q2	2017Q3	2017Q4
T+30 days	0.3628	0.2449	0.3954	0.4263
T+45 days	0.3624	0.2449	0.3547	0.3527
T+2 months	0.3999	0.3641	0.3824	0.4353

Table 2: Year-on-year growth rate (%)

	Period			
	2017Q1	2017Q2	2017Q3	2017Q4
T+30 days	1.5704	1.4946	1.5598	1.5598
T+45 days	1.5700	1.4946	1.5036	1.4853
T+2 months	1.6476	1.6291	1.5753	1.6228

(17) https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/the-annual-and-quarterly-working-time-accounts

Germany

1. Sources and their availability/timeliness

Currently, around 60 individual statistical sources obtained through different reporting channels are evaluated for the calculations at the various stages. Most of them are specialised official statistics for various sub-sectors of the economy (agriculture and forestry, fisheries, industry, services). Alternatively, they may be other branch-specific employment information reported monthly, quarterly or annually by enterprises and their establishments (for example in the field of postal services/telecommunications, railways or the financial system). The sources differ in their periodicity and coverage.

The main sources of employment data are:

- monthly administrative data from the social security system complied by the Federal Employment Agency (employee statistics and statistics on marginal part-time employment);
- labour force survey;
- public service personnel.

Data timeliness:

- employee statistics/marginal part-time employment: provisional data for the month preceding the reporting month (t+60 days - preliminary; t+180 days - final);
- labour force survey: data are available 3-4 weeks after the end of the reporting month (t+20 days);
- public service personnel: as of 30 June of the respective year; the results usually published one year later (t+360 days);
- other source data: timeliness varies between a few weeks and several years.

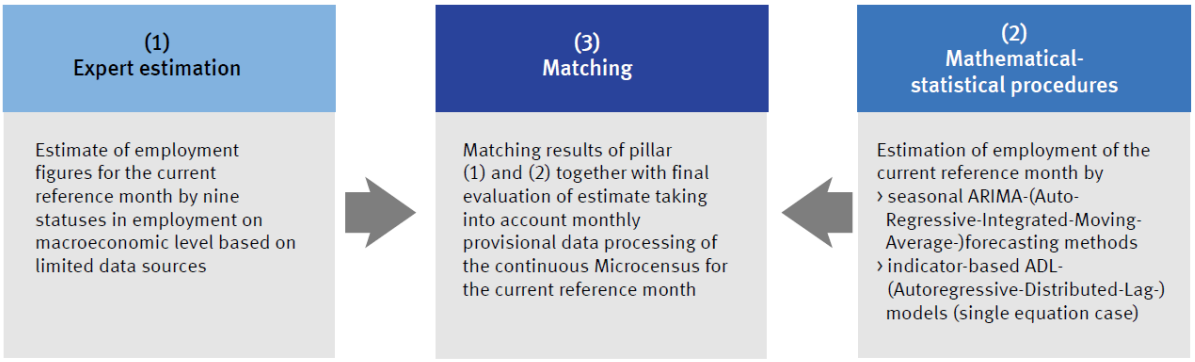
2. Variables to be estimated

As part of the labour market statistics for Germany, the Federal Statistical Office publishes the overall monthly employment figures according to the national concept as unadjusted figures and seasonally adjusted figures, together with unemployment figures and rates, both derived from the labour force survey, around 30 days after the end of the relevant reference month. The monthly employment data based on the domestic concept are also available at this early point in time. A first quarterly employment figure for the overall economy is available on the date when the result for the third month of any quarter is released – i.e. about 30 days after the reference month. Quarterly results on self-employed persons, including family workers and employees, by ten groups of industries (plus manufacturing) According to the Statistical Classification of Economic Activities in the European Community, 2011 edition (NACE rev. 2) are regularly published 45 days after the end of the reference quarter. In the first calendar week of any year, the Federal Statistical Office also publishes a first provisional employment estimate for the previous reference year.

3. Description of the estimation process

The German monthly flash estimation for employment is based on three different pillars. The main pillar is comprised of source-based expert calculations, carried out on a differentiated basis according to the status in employment, but exclusively at aggregated national level. As these calculations and estimates are based on a still very incomplete data basis, additional mathematical and statistical forecasting procedures are used, which can be carried out in the short term and independently of the inclusion of employment statistics data sources. The results determined independently of each other according to the two methods are subsequently evaluated against the backdrop of general economic performance and merged into a final result. At the same time, the monthly results of the labour force survey and further macroeconomic labour market indicators are included. This mix of various methods ensures a high degree of flexibility and utilises potential synergy effects by means of comparisons. The time series analysis and the regression approaches thus provide important support in estimating the prevailing trends and seasonal movement in employment. By contrast, peculiarities and irregular effects, for example the influence of weather factors or

new statutory rules with employment policy-related effects, are taken into account in the expert calculations.



4. Use of modelling

Different models are used. Usually the employment time series show trend characteristics, random walk processes and seasonal and non-seasonal patterns. It is therefore recommended to use the ARIMA model. The Germany flash estimate for employment uses ARIMA model selections, which are based on the Box-Jenkins procedure. However, sudden changes in the economy are difficult to capture accurately by time series analysis. Therefore, the use of regression analyses is also recommended. Earlier indicators relating to employment are needed. Germany uses ADL models, utilising data on unemployment, new orders, the production index in the manufacturing industry and business survey results.

In the German national accounts, seasonal and calendar adjustment (i.e. adjusting the original, unadjusted figures for seasonal and calendar variations) are based on the internationally preferred Census X-12-ARIMA method (in co-operation with the Deutsche Bundesbank).

5. Results, publication, revisions

All new employment accounts results (monthly, quarterly, yearly) are published as a press release and are transmitted to Eurostat. The data are always available as a time series at: http://www.destatis.de.

Regular revisions are minor corrections for individual quarters or years. They are performed in the course of current calculations and can generally occur during any release date. Such revisions are performed to include in the system current information that differs significantly from the data bases previously available. In this way, data users are supplied with the best possible results for analyses and forecasts.

Usually, the data for the quarters of the current year are checked at every quarterly release date, while data for the last four years, including the relevant quarters, are revised only once a year (in August).

Major and fundamental revisions of national accounts take place approximately every five years (latest revisions in 2014, 2011, 2005 and 1999).

Reasons for comprehensive revisions may include:

- introducing new concepts, definitions or classifications into the accounting system;
- integrating new statistical bases for the calculation that have not yet been applied;
- applying new calculation methods;
- modernising the presentation and, where required, introducing new terms;
- enhancing international comparability.

The most recent major revision of the national accounts took place in 2014 when the new European system of national and regional accounts (ESA 2010) was implemented throughout Europe. The next major revision will take place in summer 2019 and will be published in August 2019.

Estonia

1. Sources and their availability/timeliness

The Estonian labour force survey (ELFS) is used as the data source.

Statistics Estonia does not produce employment flash estimates. National employment data based on the ELFS are published at t+45 days and NA employment data based on the ELFS are also available at t+45 days.

2. Variables to be estimated

Statistics Estonia estimates the number of employed persons, employees and self-employed persons by the A10 breakdown (domestic concept) at t+45. The time series are also seasonally adjusted.

3. Description of the estimation process

The ELFS data are adjusted as much as possible to NA definitions, i.e. conscripts are included in employment data and residents working outside the economic territory (abroad or in extra-territorial organisations in Estonia) are excluded from employment data.

4. Use of modelling

For seasonal adjustment, JDemetra+ with the TRAMO/SEATS method is used (no calendar or working day adjustment). For the seasonal adjusted totals the direct method is used.

5. Results, publication, revisions

Data are final at t+45, with no revision. National employment data based on the ELFS are published at t+45 days and NA employment data are published at t+2 months.

Spain

1. Sources and their availability/timeliness

The main source used is the labour force survey. Its periodicity is quarterly and it is published between t+23 and t+30 days after the end of the reference period, .i.e. in the fourth week of January, April, July and October.

Other auxiliary sources for flash estimates of employment are:

- statistics on social security affiliation: Statistics on Social Security Affiliation of Workers, elaborated by the Ministry of Labour, Migrations and Social Security. This is not a survey, but an administrative record. Its periodicity is monthly, with a minimum lag of t+3 or t+4 days after the end of the reference period. It gives information about jobs;
- short-term statistics: Retail Trade Indices (monthly, t+30) and Services Sector Activity Indicators (monthly, t+51), elaborated by the Spanish Statistical Office (INE). They give information about the evolution of jobs in these sectors;
- tax sources: Sales, Employment and Salaries of Big Enterprises, elaborated by the Tax Agency. Big enterprises are considered to be those with a turnover of more than EUR 6 million in the last calendar year (monthly, t+45). It gives information on employee jobs within this group of enterprises.

2. Variables to be estimated

Persons employed: all persons engaged in productive activity that fall within the production boundary of the national accounts.

Hours worked: the aggregate number of hours actually worked as an employee or self-employed person during the accounting period, when their output is within the production boundary.

Jobs: defined as an explicit or implicit contract between a person and a resident institutional unit to perform work in return for compensation for a defined period or until further notice.

Full-time equivalent jobs: defined as total hours worked divided by the number of hours worked in full-time jobs within the economic territory.

3. Description of the estimation process

Firstly, a quarterly indicator for each of the variables of point 2 and for each industry is constructed. These are produced using the available data from the LFS regarding activity, professional situation, main or secondary job, type of working day, hours worked, etc. The concepts of employment in this survey are adapted to the national accounts concepts as far as possible.

For each estimated variable, a linear regression with auto-regressive errors is made between the year-on-year rates of the quarterly indicators and the quarterly published data. As a result, the preliminary quarterly aggregates for the four variables are obtained.

In the case of the general government sector a contrast is made between the results and the available information (referred to the first or the second months of the quarter, depending on the subsector) on compensation of employees from the government finance statistics.

Finally, a consistency check is made between the results obtained and the rest of the available employment sources as well as with the estimation of the rest of the variables, mainly value added through productivity.

4. Use of modelling

Predictions are made for the missing months of some auxiliary sources using ARIMA models.

5. Results, publication, revisions

Since the second half of 2018, advanced estimates results of the quarterly main aggregates of the total economy are published at t+30. In this publication, the results of the flash estimates of employment are

included. Two months later, at t+90, the revised final data are published. The main reasons for revisions between t+30 and the final estimation at t+90 would be the complete availability of the sources of employment and the availability of the final estimations of other key variables, such as valued added and employee compensation, which are needed to carry out consistency checks. In the third quarter there is additional reason for revision as the annual data are incorporated in September (t+90).

Croatia

1. Sources and their availability/timeliness

The main source for employment flash estimates is data based on the labour force survey (LFS) conducted by the Croatian Bureau of Statistics (CBS) with input from the regional units (counties) and the administrative body of the City of Zagreb authorised for official statistics activities. The goal of the survey is to measure changes in size, structure and characteristics of the active and inactive population. The survey is carried out continuously with quarterly data processing resulting in equality in estimation between t+30 and t+45 due to the sparsity of revisions. The source data is available early enough to make estimations at both t+30 and t+45 days.

2. Variables to be estimated

Employment flash estimates have been calculated since October 2017 when the first employment estimates (i.e. first three quarters of 2017) were sent to Eurostat. Estimation outcome depends on previous values, current and past errors (white noise) and outliers set as additional regression variables. Depending on time allowance and resources, the CBS is considering developing extra forecasting models such as Vector Autoregressive Model (VAR), in order to obtain other results that might serve for comparison with the results of the existing model.

3. Description of the estimation process

The first step in the analysis was to determine the stationarity of time series using the programming language R with a view to analysing the autocorrelation function and applying the Dickey-Fuller test for stationarity. By looking at the autocorrelation function, it was noted that lags are enveloped with a decreasing linear function, which indicates that our time series is not mean-stationary. Furthermore, it was suspected that the time series was not variance-stationary either, which was later confirmed by the automatic procedure in JDemetra+. In order to make the time series stationary, certain transformations were applied. Since first differences might appear as negative values, log transformations are made first. Following that, first differences are calculated on the log-transformed data with this option chosen in JDemetra+ when compiling flash estimates. Besides stationarity, irregularities occur in the form of outliers and trend component on the one hand and regularities in the form of seasonality on the other. As suggested in the Overview of GDP flash estimation methods handbook, an automatic procedure for detecting outliers is applied. For the first results, seasonal ARIMA models have been used.

4. Use of modelling

Estimating flash estimates of quarterly LFS employment data means estimating recent quarters, when the data for computing is not yet available. As previously mentioned, a seasonal ARIMA, with outliers taken into account, has been used to compile employment flash estimates. For each quarter in the testing phase, many seasonal ARIMA combinations are examined and analysed. The ARIMA model that best describes a forecast depends on several criteria, such as mean error, which should be very close to zero, and low root mean square error. The model that best fits the estimated value is chosen among the five available, having regard to the lowest Akaike information criteria. The seasonal adjustment of the flash estimates at both t+30 and t+45 was done in the same way as for data at t+70 using JDemetra+.

5. Results, publication, revisions

Employment data based on the labour force survey is regularly published at quarterly level. Unlike the labour force survey, the results of employment flash estimates are currently in the test phase and are not published. The estimates have been sent to Eurostat at t+30 and t+45 days as confidential data.

Italy

1. Sources and their availability/timeliness

SHORT-TERM INDICATOR USED FOR EMPLOYEES AND SELF-EMPLOYED PERSONS (T+60; T+45; T+30)

The labour force survey (LFS) represents the main source for the quarterly estimation of resident employed persons. LFS data are a good proxy for NA employment (both for employees and self-employed persons) in total economy and in service activities.

Quarterly time series are available for internal use at t+40 days for employees and self-employed persons (with or without contract) in ten sectors of economic activity.

Provisional monthly indicators are available at t+28 days. These include persons, job and hours worked for four economics sectors (agriculture, forestry and fishing; industry (NACE B-E); construction; and total services) for employees and self-employed persons. These series are revised and become almost definitive at t+40.

The LFS covers a considerable part of NA employment, in particular registered first jobs and a share of unregistered employed persons (i.e. those who are not registered by the company or not visible to the welfare and tax institutions).

Quarterly data are only revised in seasonally adjusted data. The reason for this is that seasonal factors may be revised due to new information and/or increasing sample size.

SHORT-TERM INDICATOR USED FOR EMPLOYEES (T+60; T+45 ONLY FORECASTING FOR INDUSTRY SECTOR)

Business indicators (OROS Survey –Employment earnings and social security contributions). Oros indicators are estimated by the integration of Social Security data (employers' social contribution declarations to Inps, the Italian Social Security Institution) and monthly Large firms Survey data (LES).

These indicators consider enterprises and private institutions established in Italy with employees that in the reference quarter paid taxable wages and salaries for social security purposes, in the industry and services sectors (NACE rev. 2, sections B-S, excluding O). Each quarter is available for internal use at t+58 days. The OROS survey releases a provisional estimate for the current quarter and a final estimate for the same quarter of the preceding year. The provisional data can be revised for the next three quarters until the final estimate is released after 12 months. The discrepancy between the preliminary estimate and the final estimate depends on many different factors[18].

SOURCE USED FOR EMPLOYEES IN 'ADMINISTRATION AND OTHER PUBLIC SERVICES SECTOR'

Annual forecasts produced by the government in the economic and financial planning document (available in May) are used to estimate the latest year. A trend is then used to produce quarterly data.

2. Variables to be estimated

The labour input that contributes to the economic output during the reference period is measured using four different definitions of employment available at t+60: domestic employment (residents and non-residents employed in domestic production units); jobs; full time equivalent (FTE) (jobs converted into standard full-time equivalents); and hours actually worked by all workers (total hours worked).

[18] Examples: a) improved quality and completeness of the final version of the administrative microdata compared to the preliminary version; b) the annual revision of the large enterprises survey (LES) data referred to the previous year, included in the OROS estimates delivered in the first quarter; c) updating of structural variables based on other external sources (e.g. NACE rev. 2 classification).

Quarterly labour input series are compiled following the ESA 2010 definitions and structure (P10 classification). Data are consistent starting from the first quarter of 1995, and are released at t+60 days after the reference period. They are expressed in raw form, in calendar adjusted form (in the case of the hours worked series), and in seasonal and calendar adjusted form. The series are available for employees and self-employed persons.

A preliminary estimate of total employment is produced both at t+30 and at t+45 days after the reference period. These estimates are expressed in raw form and in seasonally adjusted form19. Data are transmitted to Eurostat in confidential form (total employment in raw and seasonal form) in order to produce the European flash estimate. They are not published nationally.

3. Description of the estimation process

Preliminary series of total employment data produced at t+45 days after the reference period are estimated using the methodology (same econometric models) and almost all sources used for regular quarterly estimates, even if estimation of the number of employees in the industry sector (not including construction) is made using forecasting techniques to estimate the latest quarter. Total employment is the sum of four economic sectors obtained using different regressions: four for employees and ten for self-employed persons. The services sector for self-employed persons is obtained as the sum of the output of seven services sectors regressions. For employees, the output of total service regression represents a constraint for the sum of sectors 4-10 (P10 classification, ESA 2010).

Starting from the third quarter of 2016, another preliminary estimate of total employment is produced at t+30 days after the reference period since monthly LFS indicators (provisional data) are available at t+28 days. From these monthly data, a quarterly average can be computed (weighted with weights that correspond to the number of weeks (4 or 5) in each month).

4. Use of modelling

Estimates of quarterly labour input data are the result of integrating and comparing different statistical sources and indirect methods of estimation using related time series. The indirect approach with related series permits the extrapolation of the quarterly profile for the most recent year (not yet available from the annual accounts). The quarterly series are obtained through temporal disaggregation of the annual series by means of related short term indicators. The basic assumption is that the quarterly pattern (unknown) of the annual series of NA is approximately the same as that of the related indicators.

5. Results, publication, revisions

Data are transmitted to Eurostat in confidential form (total employment in raw and seasonal form) at t+30 and t+45 days after the reference period, in order to produce European flash estimates. They are not published nationally.

This estimate is revised at t+60 days after the reference period in the case of employees in the industrial sector when business indicators (OROS) become available. New releases revise the data of the current year and the preceding 4 years. The data become final 4 or 5 years following their first release. When a major revision occurs, the entire time series are revised. Quarterly figures can be revised for three main reasons: (1) changes in the basic data source; (2) revision of seasonal factors due to new information and/or increasing sample size; and (3) changes in the annual data (in February when the fourth quarter is produced and in October when a revision of second quarter is produced).

Lithuania

1. Sources and their availability/timeliness

At t+30 days, flash LFS data (monthly frequency) on employed persons aged 15-74 are available.

At t+45 days, LFS data (quarterly frequency) on total employment are available.

2. Variables to be estimated

Total employment in persons (no separation between employees and self-employed persons) is estimated for t+30 days.

Total employment at A10 breakdown (no separation between employees and self-employed persons) is estimated for t+45 days. Variables at t+45 days are produced: total employment in persons and total employment in hours.

3. Description of the estimation process

The general formula for NA employment at t+30 days is as follows:

NA employment = employed persons (LFS) + compulsory military service – temporary workers abroad

Employed persons (LFS) consist of two components: persons aged 15-74 and persons aged 75 and older. Only the first component is available.

Data for three indicators (data on compulsory military service, temporary workers abroad and employed persons (LFS) aged 75 and older are estimated using ARIMA models. Changes compared to the previous quarter and to the same quarter of the previous year are analysed. Expert analysis and corrections are carried out.

The estimation process at t+45 days consists of two steps:

1. ARIMA models are used to evaluate the estimates. ARIMA models are applied separately for each of the series (employees and self-employed persons).
2. The estimates of employees and self-employed persons are aggregated. Changes compared to the previous quarter and to the same quarter of the previous year are analysed. Expert analysis and corrections according to LFS data are carried out.

4. Use of modelling

JDemetra+ (version 1.0.4.) is used to model employment estimates at t+30 and t+45 days after the end of the quarter. Models and parameters are fixed for one year when Q4 data are available. If necessary, models and parameters can be revised more often. Models are chosen using a detailed analysis module. Logarithm transformation is tested. Seasonal and calendar factors are estimated, two regressors (working days and weekend days) are used, outliers are detected and corrected, and the Lithuanian national holiday calendar is used. The economic sense of the components of the model are analysed.

Traditional statistical methods are used to validate the models. Diagnostic statistics are used to check the quality of the models. The significance levels, number of outliers and statistics on residuals in data series are checked.

5. Results, publication, revisions

The estimates at t+30 days are not published nationally. Only test estimates have been transmitted to Eurostat as of 2017Q1. The t+30 test estimates, t+45 flash estimates, and final estimates of employment are analysed on an ongoing basis.

The estimates at t+45 days are not published nationally and are only transmitted to Eurostat. The flash estimates at t+45 days and final estimates of employment are analysed on an ongoing basis. To evaluate the accuracy, mean absolute percentage error (MAPE) is used. The estimated error (2015-2016) of total employment in persons is 0.6 per cent, while for total employment in hours it is 1.5 per cent.

Hungary

1. Sources and their availability/timeliness

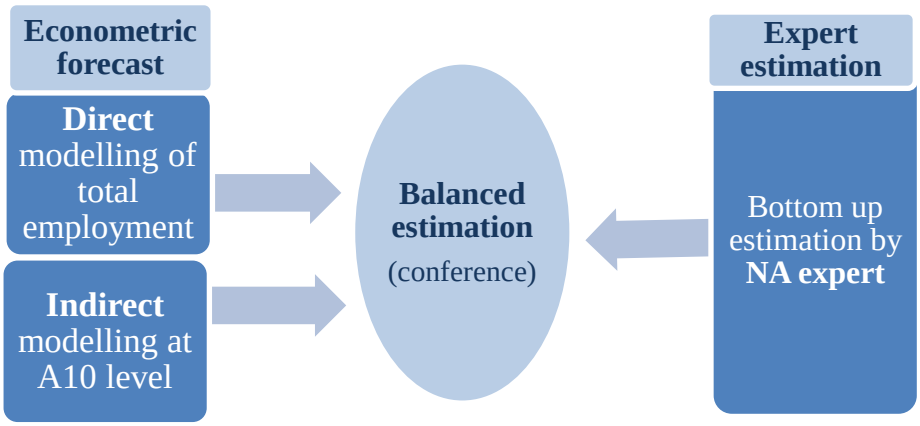
The main data source for flash estimates of quarterly employment according to the NA is the labour force survey. Quarterly data are available at t+35 days.

2. Variables to be estimated

The sum of total employment is estimated quarterly at t+45 days.

3. Description of the estimation process

The estimation process contains three steps as shown in the following illustration:



An econometric model is set up directly to estimate total employment, and indirectly at A10 level. Secondly, a previous expert estimation is also produced at NACE 2-digit level (as is done for t+60 days). From this estimation, an index method is used to extrapolate the time series. The main data source is the LFS, but quarterly data from the monthly (institutional) labour survey (relating to STS) is also used. The results of these two approaches are compared in a balancing estimation meeting where the experts of each field discuss the results and make the best decision.

4. Use of modelling

The econometric model is set up directly using total employment minus residents working abroad from the labour force survey as an explanatory variable. The least squares ARIMA model is applied using EViews 8.0 software. Quarterly GDP with a lag of 4 quarters is also examined as an explanatory variable in each quarter. In some quarters, it shows significance while in others it does not. Bottom-up modelling is also carried out at A10 level using total employment in the correspondent branch minus residents working abroad from the labour force survey as an explanatory variable. The GDP of the correspondent branch and its lagged version are also examined in each quarter.

5. Results, publication, revisions

Hungary does not publish employment data according to the NA at t+45 days. This is still in an experimental stage. Work in this field began through the participation in the Eurostat task force on the same subject. The results are sent to Eurostat confidentially as test results. So far, Hungary has no plans to publish such data.

Austria

1. Sources and their availability/timeliness

The main source used is data from the social security system. After t+30 days, there are preliminary results available for the last quarter. Final results are available after 3 years, but only minor revisions are made when new data are transmitted by the social security institution.

For several subgroups of self-employed persons some estimations are produced. For unpaid family workers the source is the LFS. Preliminary data are available after t+30 days, while final data are available after t+80 days.

Information on self-employed cross-border inbound commuters, which are in general involved in the security data set, are available when the annual figures are released (t+9 months).

2. Variables to be estimated

Austria estimates the number of self-employed persons and employees by the breakdown A10 (domestic concept) at t+45. The time series are also seasonally adjusted.

3. Description of the estimation process

The social security institution provides microdata on personal level with a unique ID. Every person who is employed or self-employed in Austria has to be registered. In the event that a person has more than one job, an automated process is carried out on the microdata, in order to obtain the main job. In principle, we use information on the ‘type’ of work, the unique personal ID and the period of insurance to obtain the main job.

For the preliminary LFS, in general weights from the last quarter are used. To subtract cross-border inbound commuters, information from the previous year and a linear trend are used. The time series are also provided in seasonally adjusted form.

4. Use of modelling

In order to make the seasonal adjustment, JDemetra+ is used using the TRAMO/SEATS method (no calendar or working day adjustment). For the seasonally adjusted totals, the indirect method is used.

5. Results, publication, revisions

The results are published nationally after t+45 and are forwarded to Eurostat, with the last quarter revised. There are no revisions at t+60 as no further information is available. When the annual figures with more detailed information for the NACE breakdown are derived, the quarters are adjusted. Although only minor revisions for the totals are made, the NACE revisions are greater.

Poland

1. Sources and their availability/timeliness

For the needs of the ESA 2010, employment data are compiled mainly on the basis of the labour force survey and using information reported by enterprises (annual national Z-06 report).

Preliminary LFS results are available after t+40 days. Preliminary results of the Z-06 survey are available after t+5 months. As a result, before the data become available information from the previous year is used.

2. Variables to be estimated

Poland estimates the employment, employees and self-employed (domestic concept) at t+44. The time series are also seasonally adjusted. Flash estimates which are transmitted cover only totals and in the regular transmission at t+60 complete data with the A10 breakdown are sent.

3. Description of the estimation process

The data from the LFS have been adjusted by excluding residents working abroad and including foreigners working for Polish employers. The data on foreigners working in Poland are taken from the annual survey on employment, wages and salaries, and hours worked, carried out by Statistics Poland on the Z-06 report form.

4. Use of modelling

Seasonal adjustment (without working days adjustment) is performed using the TRAMO-SEATS method applied in the JDemetra+ software tool. The indirect method is used for aggregates. Adjusted aggregates are calculated as the sum of adjusted components (sections aggregated to the total, employees and self-employed persons aggregated to employment).

5. Results, publication, revisions

The results are compiled for the needs of the ESA 2010 and are forwarded to Eurostat within t+44 days (only total data). In regular transmission, t+60 days data are not changed and are completed by A10 breakdown. Poland does not publish these data nationally. The data for four quarters are usually revised in December of the following year when the final regional employment data are compiled.

6. Other information

More information may be found in the QNA inventory for Poland available at:

http://ec.europa.eu/eurostat/web/national-accounts/methodology/member-states-accounts/qna-inventories.

Slovakia

1. Sources and their availability/timeliness

The main sources used to compile employment flash estimates in Slovakia are the LFS, business statistical surveys (quarterly) and administrative sources. Quarterly statistical surveys are organised by the Statistical Office of the Slovak Republic in line with the Law on State Statistics. This group covers surveys specifically focused on employment data (for general government and NPISH units) and surveys focused on production data with a separate module for employment (for financial and non-financial corporations). Administrative sources (monthly data) are used to adjust the LFS data to the ESA 2010 domestic concept and to estimate employment for the units not covered by business surveys (self-employed persons, employers and their employees from the households sector). The main institutions serving as administrative sources are the Social Insurance Agency, the Office of Labour, Social Affairs and Family of the Slovak Republic and the Statistical Office of the Slovak Republic.

LFS and business survey data are available at around t+37 days after the reference quarter but no sooner than t+30. Administrative data relevant for t+45 flash estimates are mostly available only for the first two months of the quarter.

2. Variables to be estimated

Variables of total employment in persons, employees and self-employed persons (and thus also QoQ and YoY percentage changes) are estimated at t+45. Total employment is also available in seasonally adjusted form. Then, at t+60, the total employment, employees and self-employed in persons and in hours worked are estimated together with the breakdowns of all variables by NACE sections and seasonal adjustment.

3. Description of the estimation process

Total employment according to the national accounts ESA 2010 methodology is compiled by balancing the 'supply' of labour with 'demand' for labour. Supply of labour is represented by data obtained through the LFS, which are adjusted with regard to the coverage and domestic concept of the ESA 2010. The demand for labour is represented by data from statistical business surveys, which are already expressed in the domestic concept. They are combined with administrative sources and expert estimates.

Adjustments on the supply side start with the total number of persons working in the main job recorded in the LFS. This number includes persons who are officially registered as unemployed but who take part in so-called activation schemes and receive remuneration in the form of social benefits (within the framework of active measures taken by the Labour Office). These persons are not included on the demand side (i.e. in business surveys) and therefore their number is deducted from the total number of persons working.

Adjustments to the domestic concept include deducting the number of persons working abroad (less than 1 year) and the employees of non-resident institutions in the Slovak Republic (LFS figures) from the LFS data. Non-residents working for resident units (less than 1 year), employees of Slovak embassies and members of the Slovak armed forces located abroad are added. Adjustments regarding coverage include adding the estimated number of working persons living in collective households (students living in dormitories, church workers, members of the clergy, etc.).

The data from the business statistical surveys (demand side) already correspond to the domestic concept and therefore no further adjustments are needed. The numbers of employees are estimated for particular institutional sectors using data from surveys or administrative sources. The number of persons on maternity leave (taken from the LFS) is added as this is not included in the statistical surveys. The number of self-employed persons being estimated using administrative sources (social insurance, statistical register of the Statistical Office of the Slovak Republic) is added and compared with the number of self-employed persons recorded in the LFS.

Subsequently, the adjusted figures from both sides are compared and balanced. The comparison of longer time series shows that employment from the supply side (LFS) is always higher than from the demand side (surveys). It is assumed that employment data taken from the LFS is better at covering short-term, seasonal jobs and also includes part of the non-observed economy. As a result, employment from the supply side is adjusted within the tolerance of statistical deviation and the remaining difference is

considered as non-registered employment and added to the number of employees in the households sector on the demand side. The resulting figure is treated as a definitive estimate of total employment for the entire economy.

4. Use of modelling

When compiling t+45 employment flash estimates as described above, no modelling methods are used since enough direct sources are available at t+37. In the case of administrative sources, when the third month of the quarter is missing, the average YoY growth rates of the first two months are used to produce estimations.

5. Results, publication, revisions

At t+45, flash estimates of total employment are published on the web portal of the Statistical Office of the Slovak Republic (SOSR), together with YoY growth rates from unadjusted and seasonally adjusted figures and QoQ growth rates from seasonally adjusted figures. The time series of total employment with the flash estimate is also transmitted to Eurostat at t+45. The detailed breakdowns by NACE A10 industries for variables of employment in persons and hours worked are forwarded to Eurostat at t+60 and then published in SOSR databases around t+68.

The figure of total employment is not usually revised between t+45 and t+60 because no more relevant data sources become available after t+45. The major revisions of employment data occur in September (annual data) and in November (quarterly data up to the second quarter of the current year).

6. Other information

When producing the NACE breakdowns of employment figures, statistical survey data are mostly used. The major revisions of the data from annual business statistical surveys are used to provide more precise information.

United Kingdom

1. Sources and their availability/timeliness

Labour force survey (LFS) data are available at (roughly) t+45.

Short-term employment survey (STES) data, which are collected from businesses, are available at (roughly) t+75.

Public sector employment (PSE) data are available at (roughly) t+75.

Her Majesty's Forces (HMF) data are available at (roughly) t+75.

2. Variables to be estimated

- persons employed (or 'workers') by A10 industry, split into employees and self-employed persons.
- hours worked by A10 industry, split into employees and self-employed persons.

3. Description of the estimation process

LFS microdata are used to calculate the number of main jobs in each (2 digit) industry, for both employees and the self-employed. A ratio is calculated, at the level of the overall economy (i.e. all-industry level), for both employees and the self-employed (respectively), of the number of main job hours to total hours. The number of employee and self-employed main jobs in each industry is then multiplied by each respective ratio to calculate the number of 'main job' workers in each industry. The remaining number of workers is allocated across industries in proportion to the numbers of second job hours worked in that industry. This is done for employees and the self-employed respectively.

For employment *status* (i.e. employee or self-employed) and industry *i*, this can be summarised as follows:

$$Workers_{status,i} = Jobs_{MainJob,status,i} * \left(\frac{Hours_{MainJob,status,i}}{Hours_{status}}\right) + \left(Jobs_{MainJob,status} - \sum_{i}\left[Jobs_{MainJob,status,i} * \left(\frac{Hours_{MainJob,status,i}}{Hours_{status}}\right)\right]\right) * \left(\frac{Hours_{SecondJob,status,i}}{Hours_{SecondJob,status}}\right)$$

The total hours worked in an industry are calculated as the sum of actual hours worked in main and second jobs worked in each industry. The average hours per worker are calculated as the total hours worked in an industry divided by total workers in an industry. Average hours per job are calculated as the total hours worked in an industry divided by main + second jobs in an industry. These data make up our 'pure-LFS' (LFSp) dataset.

A separate set of data is calculated using workplace data (i.e. STES and PSE). These data replace the LFS employee jobs data, and employee and self-employed jobs across industries are proportionally constrained to produce a total of the number of LFS whole economy jobs. Total employee hours in an industry are re-calculated as STES jobs multiplied by average employee hours per job (from the LFS), and employee and self-employed hours across industries are proportionally constrained to produce a total of the number of LFS whole economy hours. Employee workers in an industry are re-calculated as total employee hours in the industry divided by average employee hours in the industry. Employee and self-employed workers across industries are then proportionally constrained to produce a total of the number of LFS whole economy employee + self-employed workers. These data make up our 'workplace adjusted' (LFSa) dataset.

Both the pure-LFS (LFSp) and the workplace adjusted (LFSa) datasets are aggregated to A65 industries. HMF workers and hours are added to industry 84 (public administration and defence).

For the latest quarter, before workplace data are available (i.e. at t+45), workers and hours for each industry (for employees and self-employed persons respectively) are calculated by adjusting the LFSa data from four quarters ago by the year-on-year growth rate from the LFSp data. Thus, for industry *i* in quarter *t*,

the flash estimate for hours and workers is calculated as:

$$Flash_{i,t} = LFSa_{i,t-4}\frac{LFSp_{i,t}}{LFSp_{i,t-4}}$$

Data are then adjusted to account for double counting resulting from salaried working proprietors appearing in both the LFS self-employed data and workplace-based employee data.

Finally, employee and self-employed workers and hours are constrained to LFS totals. The data are then aggregated to A10 industries, seasonally adjusted, and constrained again to various seasonally adjusted aggregates from the Office for National Statistics (ONS) national labour market statistics.

4. Use of modelling

No modelling is used.

5. Results, publication, revisions

The methodology outlined in section 3 is only used to calculate workers and hours worked in the submission of workers and hours by industry and employee/self-employed status to Eurostat as part of the ESA regulations. A separate methodology is used to calculate jobs and hours in ONS publications and is outlined in the Quality and Methodology Information (QMI) accompanying the labour productivity statistical bulletin. A similar methodology to the one outlined in the QMI is used to produce hours worked for Eurostat under STS regulations, forwarded before t+90.

As of Q11994, all periods are open to revision for every data transmission, and data are forwarded to Eurostat on the same day as the ONS labour market statistics are released.

Annex II Software and programming languages for time series analysis, forecasting and seasonal adjustment

There are several commercial and non-commercial solutions for conducting forecast analysis, available for the most commonly used operating systems, i.e. Mac OS, Microsoft Windows and UNIX platforms:

- **EViews** is a software solution that provides powerful modelling tools emphasising both univariate (e.g. ARIMA) and multivariate (e.g. VARs) time series analysis. It can be found at: http://www.eviews.com/home.html. The EViews manual for time series analysis is available at:

  http://www.eviews.com/EViews8/EViews8/EViews%208 %20Users%20Guide%20I.pdf

- **R** (R Core Team, 2017) is a high-level programming language for statistical computing, data manipulation and powerful graphics (especially within a ggplot environment). It is available free of charge. Just like EViews, R can be used for univariate and multivariate time series analysis and can be found at: https://www.r-project.org/. There are many packages available for time series, which can be found at: https://cran.r-project.org/web/views/TimeSeries.html. Some packages also have vignettes, which give examples of analysis, e.g. for the VARs package (Pfaff, 2008). See:

  https://cran.r-project.org/web/packages/vars/vignettes/vars.pdf

The box below shows an example of basic R code for ARIMA forecasting:

```
#importing e.g. employment data from MS Excel named as series.xlsx
series <- read_excel("C:/Users/user_name/Desktop/series.xlsx")
view(series)

#defining the series (e.g. for the first column in series.xlsx file)
as time series, starting from first quarter of 2000
ts_data <- ts(series[,1], start = c(2000,1), frequency = 4)

#visualizing time series on a graph with title and named axis
par(mfrow = c(1,1))
plot(ts_data, xlab='Quarters', type='l', ylab = 'Employment in
thousands',  main = "Quarterly employment in Country X since the year
2000", col='red')
grid()

#plot ACF and PACF for detecting non-stationarity
par(mfrow = c(1,2))
acf(ts(ts_data), main='ACF Employment')
pacf(ts(ts_data), main='PACF Employment')

#confirm it by stationarity test (e.g. ADF)
adf.test(ts_data, alternative = c("stationary", "explosive"), k = 0)

#if necessary, make time series both mean and variance stationary
lg <- log(ts_data)
```

```
dlg <- diff(lg)

#plot the graph again to gain a first look on stationarity
par(mfrow = c(1,1))
plot(dlg, xlab='Quarters', type='l', ylab = 'Employment in
thousands',  main = "Quarterly employment in Country X since the year
2000", col='red')
grid()

#plot ACF and PACF to confirm stationarity (again)
par(mfrow = c(1,2))
acf(ts(dlg), main='ACF Employment')
pacf(ts(dlg), main='PACF Employment')

#conduct stationarity test again
adf.test(dlg, alternative = c("stationary", "explosive"), k = 0)

#choosing model based on ACF and PACF analysis (replace zeros)
fit <- arima(lg, order=c(0,0,0), seasonal=c(0,0,0))
summary(fit)

#prediction for quarter+1
result <- predict(fit, n.ahead = 1)
result

#now in the end, compute e^(result) where e denotes the base of
natural logarithm
exp(result$pred)
```

- **gretl** is a free and open-source software package for econometrics that includes regression and time series libraries with an interface in several world languages. It can be useful for both univariate and multivariate time series analysis and provides user-friendly graphs and results. It can be downloaded at:

  http://gretl.sourceforge.net/

- **Matlab and GNU Octave: Matlab** is a commercial programming language primarily intended for scientific computing in which a matrix is the main object. Since time series models are often based on matrix operations, they can also be implemented in Matlab. Helpful instructions on ARIMA models can be found at:

  https://www.mathworks.com/help/econ/arima-class.html

- Instructions on VARs can be found at:

  https://www.mathworks.com/help/econ/introduction-to-vector-autoregressive-var-models.html

  GNU Octave is also a programming language primarily intended for scientific computing but, unlike Matlab, it is available free of charge under GNU licensing rules. It contains a significantly lower number of libraries compared to Matlab. However, it is possible to write a code using standard books on econometrics as a guide. In addition to the ARIMA and VAR models, flash estimates are also possible using temporal disaggregation procedures. The Matlab and Octave code for temporal aggregation procedures is given at:

  https://uk.mathworks.com/matlabcentral/fileexchange/69800-temporal-disaggregation

  A general introduction to Matlab/Octave programming is provided by Linge and Langtangen (2016) at:

  https://link.springer.com/content/pdf/10.1007%2F978-3-319-32452-4.pdf

- **JDemetra+** is a software solution for time series analysis officially recommended by the European

Statistical System and the European System of Central Banks as a tool for making seasonal and calendar adjustments to official statistics data. The software uses the two most commonly used seasonal adjustment methods, i.e. the TRAMO/SEATS and X-13ARIMA methods. Apart from seasonal adjustments, JDemetra+ is very useful for time series analysis and forecasting, outlier detection, temporal disaggregation and benchmarking. It can be downloaded at:

https://ec.europa.eu/eurostat/cros/content/download_en

The website includes the current version, along with plug-ins for, among other things, temporal disaggregation and benchmarking. In addition, useful user guides relating to JDemetra+ can be downloaded at:

https://ec.europa.eu/eurostat/cros/content/documentation_en

- **Python**, like Matlab and Octave, is an interpreted high-level programming language used for scientific, engineering and statistical computing. Users who prefer Python over other available programming or software solutions may find some interesting links for time series analysis below:
  - For regular data manipulation, the Pandas library is essential: https://pandas.pydata.org/
  - For OLS regression and ARIMA, see: https://www.statsmodels.org/stable/index.html
  - For spectral analysis, see: https://pypi.org/project/spectrum/

A general introduction to Python programming is provided by Linge and Langtangen (2016) at:

https://link.springer.com/content/pdf/10.1007%2F978-3-319-32428-9.pdf

All links are available as of January 2019.

References

Linge, S. and H. P. Langtangen, Programming for computations – Matlab/Octave, A gentle introduction to numerical simulations with Matlab/Octave, Springer International Publishing AG Switzerland, 2016.

Linge, S. and H. P. Langtangen, Programming for computations – Python, A gentle introduction to numerical simulations with Python, Springer International Publishing AG Switzerland, 2016.

Pfaff, B., VAR, SVAR and SVEC Models: Implementation Within R Package vars, Journal of Statistical Software 27(4), 2008, available at: http://www.jstatsoft.org/v27/i04/.

R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2017, available at: https://www.R-project.org/.

Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: http://europa.eu/contact

On the phone or by e-mail
Europe Direct is a service that answers your questions about the European Union. You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: http://europa.eu/contact

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU Publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU law and related documents
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU
The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

KS-GQ-19-005-EN-C

Overview of employment flash estimation methods

The document provides a general overview of methods that are used in Member States for compiling employment flash estimates at 30 or 45 days after the quarter-end. It guides the compiler in the estimation process by following a step-by-step approach. The document discusses available source data, estimation procedures, model strategies, methods for forecasting missing data and the analysis of results. Description of countries' methodologies, links to available software and references to literature are included as well.

For more information
https://ec.europa.eu/eurostat/

Publications Office
of the European Union

ISBN 978-92-76-08006-0