European Centre for the Development of Vocational Training

Skills Forecast Methodological Framework

April 2023

Unedited Proof Copy

Note: This is not an official Cedefop publication. For any further information please contact Cedefop Expert Ilias Livanos who has been responsible for Cedefop Skills Forecast.

(ilias.LIVANOS@cedefop.europa.eu)

Table of contents

List of tables, figures and boxes

Tables

Figures

Introduction

Cedefop Skills Forecast provides comprehensive information on future labour market trends in Europe. The forecast acts as an early warning mechanism to help alleviate potential labour market imbalances and support different labour market actors in making informed decisions. This report presents the methodology underlying the forecast.

1.1. Background and rationale

Equipping the labour force with the right skills is one of the key policy focuses of the European Union's (EU) strategy for smart, sustainable and inclusive growth. Anticipation of skill needs has received more attention in the EU, particularly now with the rapid shift towards a climate neutral Europe as envisioned under the European Green Deal([1]) and REPower EU([2]) and digital transformation, as outlined by EU's digital strategy([3]), which aims at transforming how people and businesses work, while also supporting EU's 2050 goal for climate-neutrality. Thus, the European Skills Agenda (European Commission, 2020) sets out to support businesses and individuals to achieve better and relevant skills that would enable to strengthening sustainable competitiveness, as envisioned under the European Green Deal as well as ensuring social fairness by improving education access and lifelong learning opportunities and building resilience to react to crises. As a concrete goal, the European Skills Agenda, aims to strengthening skills intelligence and provide EU support for strategic national upskilling action, particularly by supporting the twin transitions.

It is in this context that Cedefop conducts regular, coherent and systematic skill demand and supply forecasts that reflect the most recent policy developments and try to anticipate the future skill requirements of such policy developments.

A variety of forecasting methods are used. Forecasting is an ongoing exercise, affected by changing reality, which means it is important to use the most up-to-date information and to reflect trends and changes to achieve the most reliable results.

([1]) https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

([2]) https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en

([3]) https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en

This publication presents the complex methodological framework used by Cedefop to forecast skills supply and demand. It does not promote Cedefop's methodology as the only correct methodology. Moreover, Cedefop's forecast does not replace those conducted at national level. Instead, this publication presents the problems that we have encountered and the solutions we have adopted to produce a unique pan- European skills supply and demand forecast.

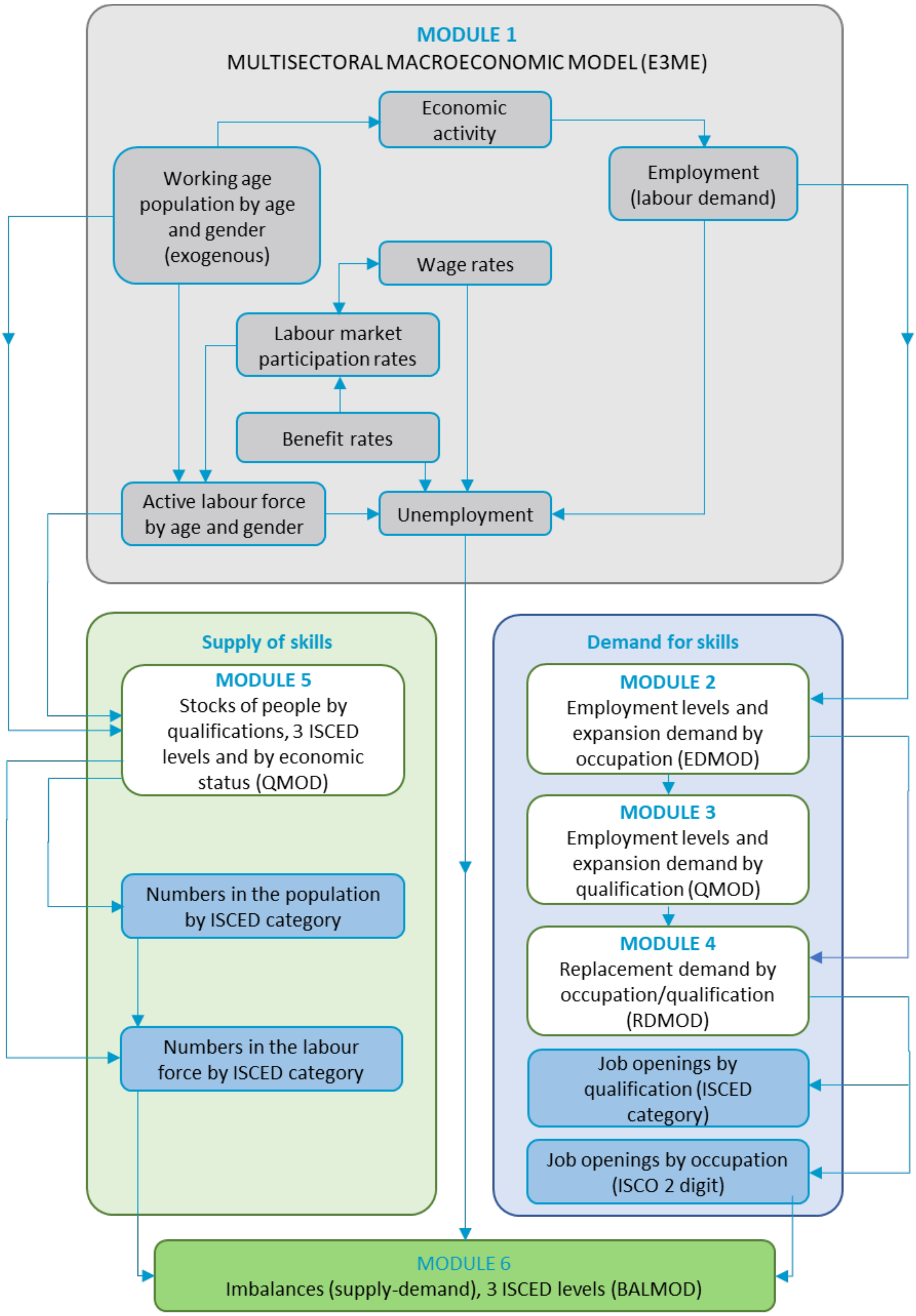
1.2. General overview of the methodological framework

Europe's pan-European forecast of skill needs requires complex methods, relying on long-term research and drawing on the expertise of several high-level European research institutions. The modelling tools have been designed to enable further development and customisation. All the details on how the general framework was developed can be found in Cedefop (2012). Forecasting is a dynamic process, and a modular approach has been adopted, which enables the different parts of the system to be improved independently. As shown in Figure 1.1, the model breaks down into different building blocks and into several interrelated components.

Even though the modelling framework has proven to be rather robust, a dialogue must be established with experts from European countries, who are likely to have much greater knowledge of recent employment trends and data sources within their own countries. By making it easy to incorporate new data and alternative or additional assumptions, the modelling framework provides an opportunity for knowledge and input of experts to be built in efficiently and transparently.

The project involved developing consistent databases and related tools to produce a comprehensive and consistent set of skill projections for all EU Member States (MSs) plus Iceland, Norway, Switzerland, the Republic of North Macedonia and Türkiye (EU-27+5). The system, models and modules rely upon official data sources, drawing primarily on Eurostat, mostly demographic data, national accounts (NA) and the EU labour force survey (EU-LFS). Compilation and harmonisation of the best possible data available for measuring employment was a major achievement of the project. Historically, most countries have invested considerable resources in developing data for their NA. In many respects estimates of employment on this basis are to be preferred as they are consistent with other key economic indicators, such as output and productivity. On the other hand, the EU-LFS has the advantage of providing measures of employment structured by skills (occupation and qualification), as well as by gender and age, which are not available from NA-based estimates.

Figure 1.1 Modelling skills supply and demand



Source: Cedefop (2023).

1.2.1. Supply of skills

The skill supply projections produce consistent pan-European projections broken down by age, gender and formal qualification([4]). The results indicate the future skill supply by highest qualification held as well as by age groups and gender for the population and labour force aged 15 and over. The skill supply projections are compatible with the skills demand projections (when focusing on qualifications).

The historical analysis and projections of overall labour supply by age and gender are provided by an extended version of the existing pan-European macroeconomic model E3ME developed by Cambridge Econometrics ([5]), which incorporates demographic and labour-supply module. E3ME models labour supply as a function of economic activity, real wage rates, unemployment and other benefit rates. At present, the model parameters are estimated for labour market participation in each country by gender and separately for different age groups. This is of key importance for modelling educational participation and attainment since these are known to be gender and age specific. This model framework is used to create a detailed set of baseline projections for labour supply, disaggregated by country, age groups and gender and covering a 10-15 year period. This model forms a key input for the analysis of the supply of qualifications and provides the link between economic activity and labour market supply. Finally, this link can be used to provide a range of projections of available skills through scenario-based analysis around the baseline forecast, indicating areas that are most sensitive to the economic climate and change.

Econometrically modelling and forecasting the supply of qualifications at the level required for this project is conceptually enormously complex at the level of detail needed. Compromises are therefore required to reduce these complexities to manageable proportions. The work has settled on a reasonably detailed and comprehensive stock-flow model, which allows an analysis at the level of three qualification levels (primary, secondary and tertiary) of 32 of EU and satellite-countries.

1.2.2. Demand for skills

The demand side involves four main elements or modules. Each module contains its own database and models. The results focus on future demand trends at a pan-European level (EU-27+5): by sector (up to 41 industries based on NACE

([4]) Skill supply is not measured by occupations as it is not possible to attribute people to different jobs after acquiring particular qualification: the occupational decisions of individuals vary significantly and cannot be predicted. Moreover, employment in occupations will change in the life course of individuals.

([5]) Detailed model description is available at: https://www.e3me.com/what/e3me/

classification); by occupation (up to 27 occupations based on ISCO classification); by qualification (three broad levels based on the ISCED classification); plus replacement demands by occupation and qualification. Together these produce estimates of the numbers of job openings (net employment change plus replacement demand) by skill (as measured by occupation and by qualification). The detailed classifications and aggregations used are provided in Annex 2.

The forecast of employment by economic sector is provided by a module which is based on results from the existing pan-European multisectoral macroeconomic model (E3ME). This model delivers a set of consistent sectoral employment projections, which are transparent in terms of the assumptions made about the main external influences on the various countries (including technological change and the impact of global competition).

E3ME combines the features of an annual short- and medium-term sectoral model, estimated by formal econometric methods, with the detail and some of the methods of the computable general equilibrium models that provide analysis of the movement of the long-term outcomes. It can also be used for dynamic policy simulation and for forecasting and projecting over the medium and long term.

The EU-LFS conducted in all countries provides a source of information for the construction of occupation-industry matrices of employment. These surveys have the advantage of being conducted regularly. They also adopt standardised sets of questions and systems of classification. While there are still some differences among countries, LFS provide a broadly consistent set of data which can be used for producing occupational employment projections within the industries identified in macroeconomic models such as E3ME. The forecasting module designed to calculate changes in employment (expansion demand) by occupation (EDMOD) based on these data works out the implications for occupational employment.

Occupational employment patterns are only one way of measuring skills. An occupational category can be understood as broadly describing a particular job (related tasks, requirements, position, etc.). Qualifications represent the characteristics of people filling these jobs as well as one of the selection criteria for filling a particular job. From the education and training policy and planning point of view, the types of qualifications typically required are important. Even with only weak data for (formal) qualifications, it has been possible to develop the module (QMOD) which allows inferences to be made about implications for qualifications.

In addition to changes in overall occupational employment levels, it is important to consider replacement demand arising from outflows from a job/ occupation, such as retirements and deaths, transition to non-employment, net migration and inter-occupational mobility. Estimating replacement demand is not straightforward and is quite sensitive to the data sources used. Ideally, detailed data on labour market outflows and transitions (mainly retirements and occupational mobility) would be required to analyse replacement demand more accurately. However, these are not currently available and therefore this forecast relies on a methodology that is based on stocks of age-cohorts by occupation and qualification, and excludes transitions from one occupation to another.

From the LFS, it is possible to analyse the demographic composition of each occupation. This allows specific rates of retirement to be estimated for each occupational class (but still not taking account of inter-occupational mobility). LFS data can also be used to estimate rates of outflow. The replacement demand model (RDMOD) has been developed on the basis of data sources that are similar to the occupational model (EDMOD). The model is driven in part by the occupational and qualification employment levels projected from EDMOD and QMOD, combined with models and information on the probability of leaving employment owing to retirement or migration and for other reasons (e.g. transition to inactivity).

1.2.3. Comparing skill supply and demand

To provide information on possible labour market imbalances and skill mismatches, a further module (BALMOD) has been added. This module compares the skill demand and skill supply projections (focusing on qualifications) and attempts to reconcile the two.

The possibility to analyse potential skill imbalances in the labour market is important from a policy and individual point of view. Such information can, in conjunction with corresponding demand estimates, shed light on possible future developments in European labour markets, highlighting potential mismatches and thus helping to inform decisions on investments in skills (especially in formal qualifications) made by individuals, organisations and policy-makers.

However, simply comparing current demand and supply projections is problematic for both practical and theoretical reasons. Although the two sets of results are based on common data and are carried out simultaneously, they do not incorporate direct interactions between supply and demand and, therefore, they cannot be directly compared. There are various other conceptual and methodological issues regarding imbalances that need to be considered to avoid misleading inferences and interpretations.

A final adjustment has been made to the estimates of employment by qualification (demand side) to take account of the labour market accounts residual. This residual measures the difference between employment as measured for the NA estimates (workplace based, jobs) and the corresponding LFS estimates (heads, residence based). Both measures are used in the project. The difference between the NA and LFS can be quite significant and needs to be considered, especially when comparing demand and supply.

Differences between skill demand and supply can include:

(a) double jobbing (some people have more than one job) or one full-time job is shared by two or more people;
(b) distinction between residence and workplace (many people do not live in the same country as they work; this is especially significant for some small countries such as Luxembourg);
(c) participants in training and related schemes who are also working in parallel (they may be included in the labour force and in education statistics – double counting);
(d) different definitions of unemployment (e.g. ILO definition versus limited to unemployment beneficiaries);
(e) statistical errors (in measures of employment, unemployment and related indicators, including sampling and measurements errors);
(f) other differences due to the use of different data sources such as treatment of nationals working abroad.

1.2.4. External validation

Any modelling framework (especially those as big and complex as the Cedefop Skills Forecast) is susceptible to produce unexpected or counterintuitive results. To identify such instances, the final results are reviewed both by the team and by Individual Country Experts (ICEs). Consultations with ICEs have a long-standing tradition in the Cedefop Skills Forecast framework. ICEs opinions and knowledge are collected through webinars, workshops and written questionnaires at different stages of the modelling process. Their feedback is incorporated into the forecast to the extent possible. ICE’s involvement starts early in the process, as soon as the first sectoral and labour supply forecast are produced. These initial results represent the backbone of the framework and getting early feedback is therefore fundamental. Moreover, the discussions with the ICEs help in building a narrative around the forecast, allows to better define the main macroeconomic trends affecting individual MSs, and ultimately strengthen the robustness and reliability of the forecast.

1.3. Structure of this report

This publication aims to provide an overview of the methodology rather than to present the results. Therefore, any results presented are only illustrative and may differ from the actual forecasting results.

The Chapter 2 of this publication presents the information on the database used in the Cedefop model. High quality data that are consistent over time and comparable across countries are a prerequisite for such a modelling exercise and for obtaining reliable and relevant results. Chapter 3 provides a detailed overview of the key features of the underlying macroeconomic model (E3ME). Chapter 4 describes the key elements of the labour supply models and attempts to introduce more of the stock-flow elements into the model. Labour demand is considered in two dimensions in Cedefop's methodology. The modelling of net changes in employment, expansion demand, is described in Chapter 5. The demand for labour created by the outflow of workers from their occupations for various reasons, replacement demand, is the scope of Chapter 6. Even though the skills supply and skills demand modules are developed under the same general framework, the comparison between these two sides of the labour market is not straightforward. Chapter 7 is hence dedicated to describing how to reconcile supply and demand as well as the imbalance indicators developed in the project. Chapter 8 summarises all the main findings of each section and outlines the next steps.

CHAPTER 2. Historical data foundation

2.1. Introduction

A coherent and consistent database containing historical time series is a prerequisite for forecasting skill supply and demand. Without data that are consistent over time and reasonably comparable across countries, the forecasts generated will have a little practical relevance, irrespective of the sophistication of the model and plausibility of the assumptions adopted.

Since the beginning of skills supply and demand forecasting at European level, a consistent database was developed comprising the key statistics used as a sound basis for the development of the model. In this chapter, the data sources used in the Cedefop Skills Forecast are briefly described. Then, issues of comparability between National Accounts data and LFS are mentioned.

2.2. Historical data sources

The historical data is comprised of mainly two types of inputs: publicly available data downloaded mainly from the Eurostat website, and LFS microdata. Table 2.1 lists the main publicly available variables and their function within the framework.

Table 2.1 Main publicly available variables entering the model

Variable (unit of measure)	Source and variable code	Breakdown	Why we need it
Employment (000s persons)	Eurostat, National accounts (nama_10_a64_e)	Sector (2-digit NACE)	Main variable projected in the forecasting exercise
Output (million EUR)	Eurostat, National accounts (nama_10_a64)	Sector (2-digit NACE)	Measure of economic activity by sector
Employees (000s)	Eurostat, National accounts (nama_10_a64_e)	Sector (2-digit NACE)	Used to calculate average wages
Compensation of employees (million EUR)	Eurostat, National accounts (nama_10_a64)	Sector (2-digit NACE)	Used to calculate average wages
Labour participation rates (%)	Eurostat, LFS (lfsa_agan)	Age groups (5-year bands)	Using participation rates and population, labour force (000s) is calculated.
Population (000s)	Eurostat (demo_pjangroup)	Age groups (5-year bands)	Used in calculating labour force

Population projections (000s)	Eurostat (proj_19np)	Age groups (5-year bands)	Used in calculating labour force projections based on the participation rate projections
Employment projections (short-term, 000s persons)	AMECO Forecast	Total	Short-term projections consistent with GDP projections Used to inform recent developments.
Compensation of employees projections (short-term, million EUR)	AMECO Forecast	Total	As above
Unemployment rate (%, historical and short term projections)	AMECO Forecast	Total	As above
Real GDP (million EUR 2010 prices, historical and short-term projections)	AMECO Forecast	Total	Used to assess short term trends
Real GDP (% growth, long-term projections)	Directorate-General for Economic and Financial Affairs - Ageing Report	Total	Long-term economy trends
Commodity prices (indices)	World Bank commodities price forecast IEA World Economic Outlook 2021	Commodities	Global commodity prices used to inform import prices

Source: The Authors.

The data shown in Table 2.1 are used to produce forecast of sectoral employment and labour supply, as explained in Chapter 3. The availability of NA historical data varies between MSs, both in terms of years and sectoral breakdowns, hence imputations were implemented to fill missing values (see Technical report (Cedefop, 2023) for more details).

The second main source of data is the EU-LFS, which is used to disaggregate the sectoral employment and labour supply forecast by occupations and qualifications, as explained in Chapter 4, Chapter 5 and Chapter 6. Table 2.2 shows the main data inputs extracted from the LFS microdata.

Table 2.2 Main data inputs from EU-LFS microdata

Module	Variable	Breakdown
Demand	Employment	Sector (NACE 1-digit); occupation (ISCO 2-digit), qualification level (3 broad levels); gender
	Employment	Sector (NACE 2-digit); age; gender
	Employment	Gender; age group; qualification level (3 broad levels); occupation (ISCO 3digit); ILOSTAT employment status (1,2,3,4)
Supply	Population	Age groups; qualification level (3 broad levels); gender
	Labour force	Age groups; qualification level (3 broad levels); gender

Source: The Authors.

The handling of the EU-LFS presents several challenges ([6]), such as sample variability and different data collection methods across MSs, and the presence of non-response in occupations and sectors. In such cases, the number of persons with 'no response' were allocated pro rata on the basis of the distribution of the missing information between the other dimensions. For example, when the information on educational attainment was not available for a particular employed person, it was imputed according to the educational structure in other dimensions, such as industry (NACE 2-digit), occupation, age group and gender.

The Covid-19 pandemic and the 2021 change in the LFS methodology caused some additional difficulties in assessing the robustness and reliability of the data, please refer to the Technical Report (Cedefop, 2023) for more detailed information on how these issues were tackled in each release.

2.3. Comparability between National Accounts and LFS data

As described in the previous sections, the two main data sources used in the Cedefop Skills Forecast are the NA and the EU-LFS. The first set of data is used to produce forecast of employment by sector and labour supply by age, while the EU-LFS is used to disaggregate the results into occupations and qualifications. However, it is important to note that the two sources present important differences.

First, the **concept of population differs somewhat**:

(a) EU-LFS statistics usually cover the population in private households, while population statistics cover the whole population, including those living in collective households (e.g. conscripts);

(b) For some MSs, the rules for defining the usual resident population differ in the EU-LFS from the rule in population statistics;

(c) Population statistics usually refer to particular dates, e.g. 1st January or mid-year for population level and characteristics. The EU-LFS statistics generally refer to the average quarterly or annual situation.

In the Cedefop Skills Forecast, population is used in the calculation of the labour force, i.e. levels are multiplied by participation rates (computed based on the LFS dataset lfsa_agan) to obtain the corresponding labour force. Hence,

([6]) See the previous methodological report for a more detailed description of these challenges (Cedefop, 2012).

differences in population between the dataset demo_pjangroup (see Table 2.1) and LFS will translate into differences in the level of the labour force.

Second, the **definition of employment also differs**:

(a) NA estimate employment according to the “domestic concept”, whereby all the persons employed in the MS are counted as employed regardless of their place of residence. The EU-LFS definition for employment covers only residents in households, which means that it does not take into account cross-border workers.

(b) The EU-LFS does not cover persons living in institutional or collective households (e.g. conscripts), unpaid apprentices and trainees and/or persons on extended parental leave, while all these categories are covered in national accounts.

(c) NA use a variety of sources to estimate employment such as business surveys, employment registers, social security registers, population census, EU-LFS and others. Although EU-LFS are among the key input in producing NA data, discrepancies in the employment levels arise due to other sources.

(d) Employment figures in the NA dataset are made consistent with other variables such as output and compensation of employees. The need to ensure consistency with other variables in the NA dataset may result in discrepancies with the EU-LFS.

(e) The EU-LFS excludes people aged less than 15, while NA do not distinguish by age.

Hence, the figures of employment and labour force in absolute levels from the two sources may differ in some MSs, and that LFS-based figures such as employment rates are not directly comparable with the NA-based figures produced by the Cedefop Skills Forecast.

CHAPTER 3. Modelling general macroeconomic and sectoral trends

The pan-European skills demand and supply forecast is based on the macroeconomic multisectoral and multi-country model E3ME developed by Cambridge Econometrics ([7]). It is designed to provide a consistent forecast of economic trends, resulting in basic labour demand and supply predictions. It is a large-scale econometric model covering all EU M Ss, plus Iceland, Norway, Switzerland, Republic of North Macedonia and Turkey, and a detailed disaggregation of 42 economic sectors, consistent with the NACE Rev2 2-digit classification (Annex 2). Interaction among economic sectors takes place through input-output relationships and links among countries are formed through international trade equations.

The first section describes the general characteristics of the model. The focus on labour market within the framework of E3ME is described in Section 3.2.

3.1. General characteristics of the model

The structure of the model is based on the ESA2010 (Eurostat, 2013) and includes detailed two-way links among European economies, energy systems and the environment. The economic system is closely tied to the model's treatment of Europe labour markets.

The econometric specification of the model makes it suitable for short and medium-term forecasting and policy analysis. E3ME is estimated and solved on an annual basis, with historical databases covering the period starting in 1970 up to the most recent year that data are available.

E3ME uses a hierarchical system of data sources, with preference given to those that use definitions which are consistent with ESA2010 and across national boundaries.

The primary source of economic data is the Eurostat NA branch. Even when Eurostat data are incomplete or believed to be of poor quality, the Eurostat definitions are adopted and the data are completed from other sources. The main example of this is the OECD Stan database, which allows additional disaggregation of some sectors. Other data sources include the European Commission's annual macroeconomic (AMECO) database, as well as information available from the World Bank and the International Monetary Fund. All time series

([7]) For more information see https://www.e3me.com/ .

data are collected on an annual basis with gaps filled in using custom algorithms developed using the Python programming language. For the older Member States (plus Norway and Switzerland) the historical database covers the period from 1970; for the 2004 accession countries data cover period from 1993; and for Romania, Bulgaria and Croatia coverage starts in 1995. The model uses permanently updated databases so the upper demarcation of the time series depends on the vintage actually used. The main cross-section data consist of input-output tables and bilateral trade matrices, which are sourced from the most recent Eurostat NA and the World Bank's World Integrated Trade Solutions database.

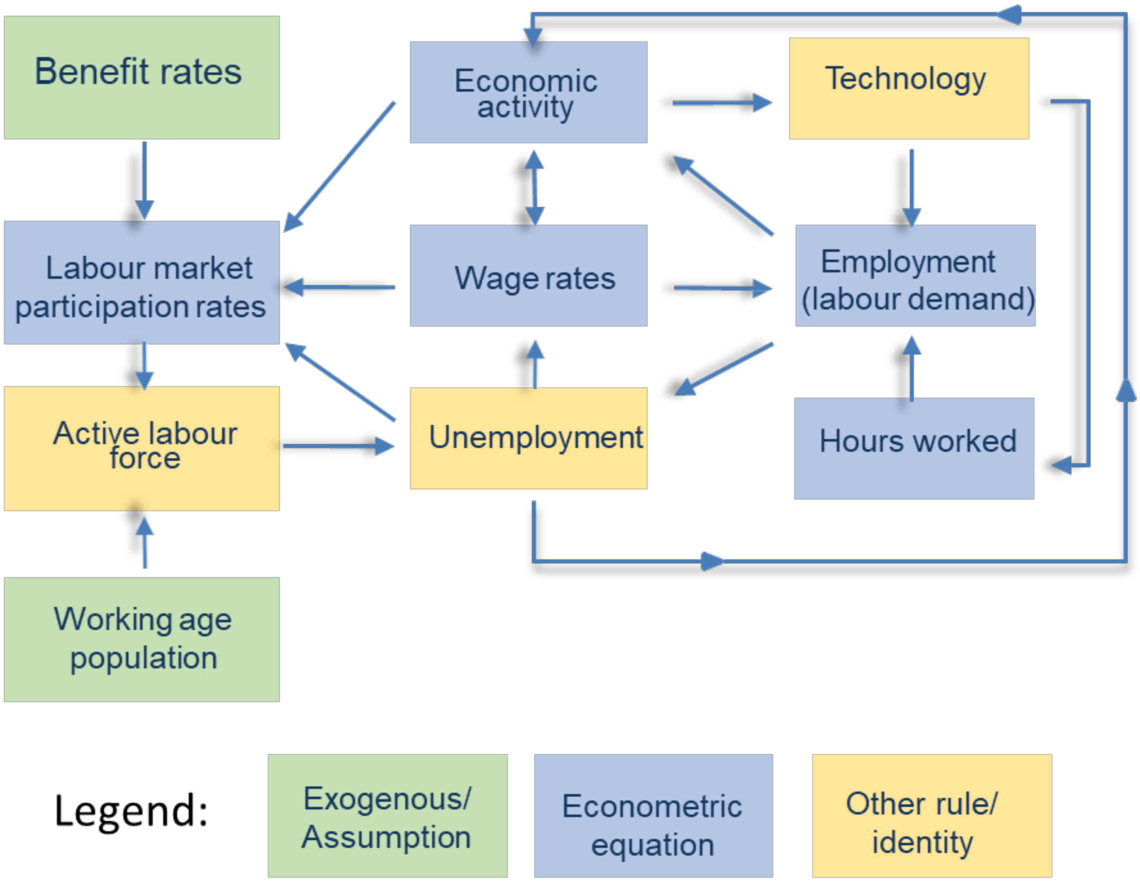
The parameters of the model are estimated empirically using formal econometric techniques. In particular, the method utilises developments in time-series econometrics, in which dynamic relationships are specified in terms of error-correction models (ECM) that allow dynamic convergence to a long-term outcome. The specific functional form of the equations is based on the econometric techniques of cointegration and error correction, particularly as promoted by Engle and Granger (1987) and Hendry et al. (1984).

3.2. Treatment of the labour market

The main role of E3ME in the forecasting project is to describe the links between the labour market and the wider economy. However, because of its detailed sectoral disaggregation, the model is also able to include a relatively complex treatment of the labour market, although it does not directly address skills and qualifications requirements or availability. E3ME labour market module (see Figure 3.1) includes equation sets for employment (as a headcount), average working hours, wage rates, and participation rates. The first three of these equations are disaggregated by economic sector while participation rates are disaggregated by gender and five-year age band.

The labour force is determined by multiplying labour market participation rates by population. Unemployment (both voluntary and involuntary) is determined by taking the difference between the labour force and employment. Employment is modelled for each industry and region as a function of industry output, wages, hours worked, technological progress and energy prices. Industrial output is assumed to have a positive effect on employment, while the effect of higher wages and longer working hours is assumed to be negative. The effects of technical progress are ambiguous, as investment may create new jobs or replace labour by automated processes; this varies among sectors.

Figure 3.1 E3ME labour market module



Source: Cambridge Econometrics.

Hours worked is defined as the average across all workers in an industry and is required because employment is modelled by headcount, rather than in person-hours. The expected related coefficient in the employment equation to be negative; if people are, on average, working longer hours, then this should have an adverse effect on job opportunities. The effect of identifying an hours-worked variable will even out when it comes to analysing productivity effects, but in countries with relatively flexible labour markets it is a good idea to try and model the effects explicitly.

The chosen model for hours worked follows the methodology of Neal and Wilson (1987). The relationship at its simplest level can be explained as function of normal hours per person per week and a measure of technological progress. The discrepancy between desired actual hours and optimal hours is assumed to arise mainly from short-run output adjustments. With a fixed capital stock, any deviation of output from its forecast level will be met largely through adjustment in hours worked, i.e., either by overtime or short-time working, while employment levels are adjusted by the firm.

The estimates for wage rates in E3ME are an important input to the employment equations in E3ME. The treatment of wage determination is based on a theory of the wage-setting decisions made by a utility-maximising union, where the union derives utility (the representative of its members) from higher real consumption wages (relative to a fallback level) and from higher levels of employment (also relative to a fallback level). The fallback level is taken to be proportional to a simple average of employment levels in the last two years in the empirical work.

In E3ME, wages are set by unions which choose wage rates to maximise utility subject to the labour-demand constraint imposed by profit-maximising firms. The form of the equation is relatively straightforward: real wages in a sector rise (with weights) if there are internal, sector-specific shocks which cause revenue per worker to rise (e.g., productivity innovations in the sector), or if employment levels are rising. Real wages are also influenced by external effects, including changes in the real wage that can be obtained in the remainder of the economy, changes in incomes received if unemployed, and changes in the unemployment rate itself. There are slight differences to this approach, in the long-run compared to short-run. The empirical evidence (e.g. Layard et al., 1991) suggests that, in the long-run, bargaining takes place over real pay, and this is imposed in the E3ME wage rate equations. However, in the short-run component of the equation for the change in wage rates can be influenced by changes in consumer prices. In addition, it has been assumed that long-run price homogeneity holds, so that the long-run economy-wide real product wage rates grow at the same rate as economy-wide labour productivity.

E3ME also includes a set of equations for labour market participation rates. The standard analysis of participation in the labour force is based around the idea of a reservation wage, such that if the market wage is greater than an individual's reservation wage, they will actively seek employment, and vice versa. It should be noted that this type of model assumes an excess demand for labour. Furthermore, in time-series studies, much of the personal background data usually used in cross-section studies to estimate preferences are unavailable, so any model is necessarily limited to variables describing human wealth (in the narrowest of senses) and market wage determination. The original variables that were available for inclusion were the market wage rate, a measure of market activity (output), a proxy for non-labour income, and some measure of the tightness of the labour market, such as the unemployment rate. The equation was later expanded to include working hours, qualifications and pension income for older age groups, as empirical analysis found these to have statistically significant impacts on participation.

E3ME includes a set of equations for labour market participation rates. When these are multiplied by the exogenous projections for the working age population we obtain a measure of labour supply. The participation rates are affected by economic output, wage rates, hours worked, unemployment, benefit and pension rates, qualifications and the ratio of service activity to manufacturing. As there is little historical precedent, changes in official retirement ages are not directly included in the equations but are taken account of exogenously. At present there is no direct link with official retirement ages. Equations are estimated and solved for each country, disaggregated by gender and by five-year age bands. The separate results for males and females indicate different patterns of activity rates between the two genders, with the activity rate for males being, in general, considerably higher than for females

E3ME does not assume market clearing in the labour market and unemployment arises when labour supply is greater than labour demand. The model database contains time series of unemployment data that are consistent with ILO definitions. Unfortunately the differences in definitions (between NA and LFS) mean that unemployment cannot be taken simply as the difference between labour supply and demand, and a residual value must be used to make the results consistent with the historical figures. This topic is elaborated more in the Chapter 7.

CHAPTER 4. Modelling qualifications supply

Although there has been a great deal of academic work on educational choice and related issues, it has only rarely focused on the development of models suitable for developing projections of the supply of skills. Some work in individual MS has usually been based on access to specific data not generally available at pan-European level.

While such studies can be valuable, they invariably are not designed to address comparative international questions or to forecast future outcomes. This means they are also not required to utilise data sources that provide consistent measures across countries, the analysis of which may require quite distinct modelling and statistical techniques. The modelling described in detail below attempts to answer the question of what the distribution of qualification levels is likely to look like 10 or 15 years forward. Note that this is distinct from *what the distribution of qualification levels will be*, because a key reason for providing the projections is to provide evidence that allows national and EU-wide interventions that will improve the actual outcomes vis a vis those projected.

The current methodology is based on analysis of changing patterns over time in the stocks of people in the population and in the labour force, defined by highest qualification held, and by country separately. These data are taken from the EU-LFS, which, despite discrepancies in coding, data availability and discontinuities due to changes in classification or definitions, etc. has emerged as a remarkably rich and consistent data set given the diversity of the countries covered. Whatever the remaining problems, the results suggest it is possible to begin to explore the implications for the future supply of qualifications in the population and the labour force. Many of the trends appear to be very robust, with certain commonalities – such as increased qualification outcomes – in all countries.

The present approach focuses on changing supply patterns, based upon a stock-flow model, without any reference to demand-side developments. In practice, the observed changes are likely to have been the result of a combination of both demand and supply influences. As a measure of skill supply, the highest formal qualification obtained by those who entered the population of working age, and the results are reported by five-year age groups and gender. In general, skills supply is determined by demographic developments, labour market participation and decisions on obtaining an educational credential. As mentioned in Chapter 3, the calculations determining the overall volume of labour supply projections are developed within the framework of the E3ME model. This chapter, however, focuses on the forecasting of qualification patterns.

Section 4.1 presents the key features of the stock model for supply by qualification as currently used. Section 4.2 describes the conceptual background of stock-flow model and how the current model is built. Finally, Section 4.3 outlines a few overarching conclusions based upon the modelling exercise.

4.1. Specification of stock model for supply by qualification

Our starting point is to consider the stocks of individuals holding different levels of qualifications, while Section 4.2 explains how these stocks are utilised in the development of a stock-flow model.

There are various theoretical considerations surrounding the modelling of skills supply, including factors that might explain changes in the qualifications structure of the labour force and of the population as a whole. However , in practice, the estimation of complex behavioural models is impractical, mainly due to data restrictions. For example:

(a) it has never proved possible to generalise the concept of "skills", as opposed to formal qualifications, in a way that would enable measures to be included in surveys such as the EU-LFS;

(b) neither does the EU-LFS do contain data on innate abilities ([8]), that are ideally required to model educational choices at the individual level;

(c) similarly, there is no information on wages so the impact of economic factors cannot be estimated easily (although new innovative match education-earnings data sets are beginning to emerge).

As a consequence, analyses that attempt a more sophisticated analysis, but lack such information, might themselves produce biased results.

A brief examination of the literature on qualification mismatches suggests that the earlier conclusions of Livanos and Wilson (2008), which reviewed the theoretical and empirical approaches to modelling and projecting the supply of people with different levels of qualifications, are still valid. Where detailed data are available, researchers have exploited them. More often than not, however, the

([8]) The data from 2020 EU-LFS Ad-hoc module are not yet available. This ad-hoc module is the first pan-European consistent survey of its kind: EUR-Lex - 32020R1642 - EN - EUR-Lex (europa.eu) paucity of information available has resulted in very simple approaches based on time series methods, which themselves have often relied on a single variable (time), rather than on multivariate, behavioural approaches.

The ideal approach would probably be to use a linked stock-flow model, with future stocks being related to past ones by an accounting relationship, including separate analyses of all relevant inflows and outflows. The latter would include flows of people into education and of those obtaining formal qualifications, as well as outflows for various reasons (including mortality, migration and flows in and out of economic activity). The present modelling exercise, for example, results in estimates of inflows and outflows over the previous ten years, as well as projected flows, which might be utilised in further analyses, which might assist in policy design at the pan-European and individual country levels.

The role of ICEs is of considerable importance for the project, since the results obtained need to be validated using local knowledge. In particular, the country results obtained from this stage of the project are typically examined by ICEs to have their feedback incorporated into the final or updated set of results.

4.2. Building a stock-flow model

Provision of information on the highest qualifications is wholly dependent on data from the EU-LFS. While longitudinal (cohort) qualification databases from which a comprehensive set of demographic accounts showing how individuals progress through the educational system and the labour market over time are beginning to emerge, they are unlikely to replace the EU-LFS for many years to come. Nevertheless, as will be demonstrated below, the EU-LFS contains a considerable amount of information on educational stocks and flows.

The model focuses on the highest level of qualification held by the individual. This imposes a hierarchical structure on education which can be checked from the results, but may not always clearly reflect the types and levels of qualifications individuals require for employment over their working lives (e.g. a degree holder may need to acquire a lower-level qualification at some stage to carry out a particular occupational activity).

The modelling work adopts the traditional threefold classification of qualifications, which are categorised as low (ISCED1-2), medium (ISCED3-4) and high (ISCED5-8) ([9]). Analysis at lower levels of the ISCED results in significant

([9]) See Section A1.3 for details on the ISCED 2011 levels of education.

problems of comparability (e.g. ISCED5) and small cell sizes (e.g. ISCED8). Even now, the low qualification group (ISCED1-2) is beginning to exhibit sample size issues, particularly during the projection period, while the growing numbers in ISCED5-8 make it a candidate for future disaggregation.

The discussion below outlines the way in which the projected shares of educational attainment by gender and age groups are constructed. Once estimated these shares are applied to the projections of the working age population for the same age and gender categories. The projections of working age population and the labour force are obtained from the E3ME model. The EU-LFS microdata are organised by year of age, rather than five-year age bands. This makes the age at which particular key educational events occur clearer (e.g. the time at which transitions from low to medium or medium to high). The results are reassembled to age bands, rather than year of age, for reporting purposes at the end of the modelling process.

The modelling activity generates forecasts of the numbers of people in the working age population by highest qualification held, broken down by age and gender. The projections of the labour force are obtained by applying the estimated activity rates to the projections of the working age population. In the final analysis, it is always ensured that the male and female results are consistent with the totals and that the three qualification levels always sum to the female, male and total population for each age group.

4.2.1. Essence of a stock-flow model

There are two main components of the stock-flow modelling: (i) the educational transitions, which focus on how individuals move through it the different levels of qualifications and, thereby, into the labour market over time; (ii) the post-formal education period in which individuals carry with them the highest qualification level they achieved at stage (i) and, thereby, how they eventually leave the labour market. Both events impact on the qualification mixes of both the population and the labour force. In most countries in the recent past, for example, the inflow of more highly qualified young people has been accompanied by the outflow of less qualified older individuals, causing the qualification level of the remaining population to rise.

While the EU-LFS does not provide information about a cohort of individuals identified at time t can be traced at time t+1, it does enable a pseudo-cohort to be constructed. In other words, a representative group of 16 year-olds in 2010 can be matched with a representative group of 17 year-olds in 2011 and so on. A hypothetical example is provided in Table 4.1, where each (pseudo-) cohort starts at age 16 and a new cohort begins in each year from 2010 to 2020. The ages at which representative individuals are observed in this example, range from 16 to 26 for the earliest cohort (starting 2010), but only the youngest individual of the most recent cohort (beginning 2020) is observed.

Table 4.1 Rolling the LFS data forward' in a stock-flow model

Year of cohort	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
2020												16
2019											16	17
2018										16	17	18
2017									16	17	18	19
2016								16	17	18	19	20
2015							16	17	18	19	20	21
2014						16	17	18	19	20	21	22
2013					16	17	18	19	20	21	22	23
2012				16	17	18	19	20	21	22	23	24
2011			16	17	18	19	20	21	22	23	24	25
2010		16	17	18	19	20	21	22	23	24	25	26
2009	16	17	18	19	20	21	22	23	24	25	26	27

Source: Authors' calculations.

If, for example, all 16-year-olds hold only low qualifications in 2010, but, say, 80% of 17 year olds hold medium qualifications by 2011 (see the two adjacent shaded cells in Table 7), then the transition rate from low to medium can be assumed to be 0.8 between those two years. If nobody in that cohort has reached the high qualification level by 2011, then it can be concluded that the proportions of low and medium at age 16 were 1,0, and 0.2, 0.8 respectively at age 17.

It should be noted that deaths and migration will affect the sizes of the stocks of individuals of different ages. Death rates tend to be very low amongst individuals of ages over which formal qualifications are normally attained. While they are clearly much more important amongst the oldest, it is only the difference in death rates between those holding different levels of qualifications that will impact on the transition proportions. The issue of migration is potentially more problematic, but, as the most recent data are for 2020, the effects of the most recent events – such as the war in Ukraine – will not be picked up.

As with earlier projections, such factors are accounted for by simply scaling the EU-LFS 'projections' up to the population forecasts provided. Such scaling of the data to projected population numbers only takes place after the main elements of the modelling – in the present case, educational mix – have been carried out.

While future work may need to take account of the impact of migration, country experts may need to consider what effects such events may have on the projected outcomes which do not fully reflect the most recent events.

4.2.2. Qualifications in a stock-flow setting

While shifting the data to align, say 16-year-olds in 2010 with 17 year olds in 2011 sounds innocuous, to cover the period over which qualification transitions are most likely to occur – which is assumed here to be between 16 and 27 – means that 16 year olds in 2009 must be aligned with 27 year olds in 2021 (again see Table 4.1).

Table 4.2 illustrates how the data are rearranged into a format used for modelling purposes. As with Table 4.1, it demonstrates what information is available for each cohort (unshaded) and what is missing (shaded). The data shown are for individuals in the low qualification group and similar tables are constructed for the corresponding medium and high qualifications groups. Separate sets of tables are developed for females, males and all individuals, for each country. The proportions across qualification levels for each gender sum to unity (e.g. all males, all females and all individuals).

Table 4.2 Projecting the transition proportions for each pseudo-cohort

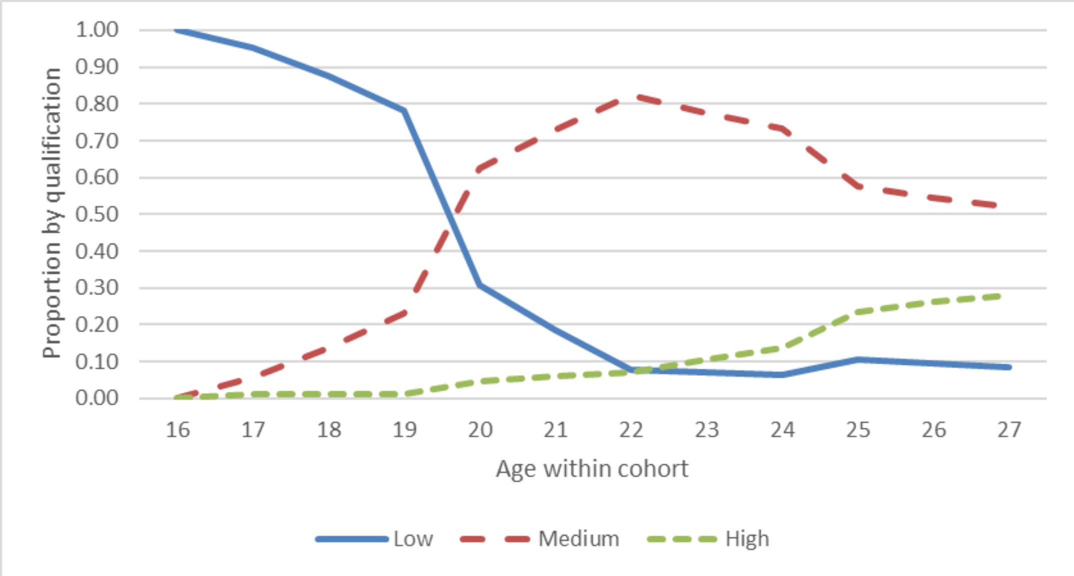
	Qualification transitions (example low)											
	Age within each cohort:											
Cohort Aged 16 in year:	16	17	18	19	20	21	22	23	24	25	26	27
2025	1.00	0.97	0.90	0.82	0.34	0.21	0.09	0.08	0.07	0.11	0.10	0.09
2024	1.00	0.96	0.90	0.81	0.33	0.20	0.09	0.08	0.07	0.11	0.10	0.09
2023	1.00	0.96	0.89	0.81	0.33	0.20	0.08	0.08	0.07	0.11	0.10	0.09
2022	1.00	0.96	0.89	0.80	0.32	0.19	0.08	0.07	0.07	0.11	0.10	0.09
2021	1.00	0.96	0.88	0.79	0.31	0.19	0.08	0.07	0.07	0.11	0.09	0.09
2020	1.00	0.95	0.87	0.78	0.31	0.19	0.08	0.07	0.06	0.10	0.09	0.08
2019	1.00	0.96	0.87	0.77	0.30	0.18	0.07	0.07	0.06	0.10	0.09	0.08
2018	1.00	0.93	0.87	0.77	0.29	0.18	0.07	0.07	0.06	0.10	0.09	0.08
2017	1.00	0.95	0.84	0.77	0.29	0.17	0.07	0.07	0.06	0.10	0.09	0.08
2016	1.00	0.93	0.86	0.73	0.27	0.17	0.07	0.06	0.06	0.10	0.09	0.08
2015	1.00	0.95	0.83	0.76	0.30	0.15	0.07	0.06	0.06	0.10	0.09	0.08
2014	1.00	0.94	0.85	0.71	0.28	0.18	0.05	0.06	0.06	0.10	0.09	0.08
2013	1.00	0.92	0.85	0.74	0.24	0.16	0.08	0.05	0.05	0.09	0.09	0.08
2012	1.00	0.94	0.81	0.73	0.27	0.13	0.06	0.07	0.04	0.09	0.09	0.08
2011	1.00	0.92	0.83	0.69	0.24	0.16	0.05	0.06	0.06	0.08	0.08	0.08
2010	1.00	0.92	0.82	0.71	0.23	0.13	0.07	0.05	0.05	0.10	0.07	0.08
2009	1.00	0.93	0.82	0.70	0.24	0.13	0.05	0.06	0.04	0.09	0.09	0.06
2008	1.00	0.93	0.82	0.69	0.22	0.14	0.05	0.05	0.06	0.09	0.08	0.08

2007	1.00	0.93	0.83	0.70	0.23	0.12	0.06	0.04	0.04	0.10	0.08	0.07
2006	1.00	0.92	0.83	0.70	0.22	0.13	0.04	0.05	0.04	0.09	0.09	0.08
2005	1.00	0.92	0.82	0.71	0.22	0.12	0.05	0.04	0.05	0.08	0.08	0.08
2004	1.00	0.95	0.81	0.69	0.23	0.12	0.04	0.04	0.04	0.08	0.07	0.08
2003	1.00	0.95	0.86	0.68	0.23	0.13	0.04	0.04	0.04	0.07	0.08	0.07
2002	1.00	0.94	0.86	0.74	0.22	0.12	0.05	0.04	0.04	0.09	0.06	0.07
2001	1.00	0.93	0.83	0.74	0.24	0.12	0.04	0.04	0.04	0.08	0.08	0.05
2000	1.00	0.94	0.82	0.70	0.24	0.13	0.04	0.04	0.04	0.08	0.08	0.07
1999	0.92	0.90	0.83	0.69	0.20	0.13	0.05	0.04	0.04	0.08	0.08	0.07
1998	0.93	0.78	0.77	0.71	0.21	0.11	0.05	0.04	0.04	0.08	0.08	0.07

Source: Research teams' calculations.

Figure 4.1 illustrates the type of results obtained for the proportions of the three qualification levels for individuals in an early cohort, for which the observed data across ages 16 to 27 are complete. The curves reflect two effects: (current) education transitions, which occur from year to year and (previous) educational trends, which affect the size (and timing) of the transitions in the past. As educational transitions are particularly important in the early years, they tend to dominate among younger age groups in the cross-sectional data and, as educational transitions become less important among older individuals, the long-term educational trends dominate the curves among the higher age groups. It is possible to just observe the as there is rising proportion of older individuals with low level qualifications and the falling proportion of those with high qualifications. This is based on fact that a smaller proportion of earlier population cohorts went on to higher levels of education.

Figure 4.1 Educational transitions and qualification proportions



Source: Research teams' calculations.

The implication of this is that the transition that occurs between age 26 and 27 is only observed for the cohort of individuals aged 16 in 2010, when many things may have changed in the intervening years that eventually alter the transition patterns that take place between age 26 and 27 for the cohort of individuals aged 16 in, say, 2020. While is clearly potentially, problematic the implication is identical for an actual (as opposed to a pseudo-) cohort as well as for simple non-cohort time series estimates. In other words, we would have needed to have traced 16-year-olds in 2010 over the subsequent 10 years before the 27/26 transition could be observed.

Educational transitions are reflected in any annual cross-sectional data for individuals, for example, the stock of individuals of working age in 2007, as shown for the EU-27 in Figure 4.1. The downward shift in low qualifications and the corresponding upward jump in medium qualifications between the ages 16 and 20 are mainly the reflection of the educational transition that occurs every year in the EU as individuals obtain their secondary education qualifications. Likewise, the downward shift in the medium qualification proportion and the upward jump in the high proportion around ages 21 to 28 reflect the educational transition that, again, broadly speaking, occurs every year in the EU as individuals obtain their tertiary qualifications. It is clear that similar patterns are likely to be present in every Member State and to differ between countries

Returning to Table 4.2, in order to proceed with a (pseudo-) cohort-based analysis, it is therefore necessary to “fill-in” the triangle of data running down to the right from 2020 age 17 to 2010 aged 27 and, thereby, provide estimates of the missing transitions. This in-fill is carried out using trends across the different cohorts for each age category – this allows a shift in transition rates at each age over time, which can differ between ages. One feature stands out immediately – that projecting transitions at younger ages will be based on more recent observations than those for older age categories. Various assumptions can be made about the manner in which the shift takes place by age, which could form the basis for simulation and scenario build exercises, but the simple assumption made here is that the trend for each year of age is linear, but can differ between ages.

This exercise in projection is extended to the future proportions shown by the years 2021 and above, represented by the shaded rectangle. In both the cases of the missing triangle and rectangle, the last ten years of data available (e.g. 2011 to 2020 for 16-year-olds and 2011 to 2010) are used as the basis for each of the cohort proportion projections. Clearly, there is a difference of over 10 years in the number of years being projected forward to meeting the needs of forecasting results through to 2035. All the projections, however, are a long way forward, with the implication that caution should be exercised the further ahead the information sought from the projections, but even more so for the older than the younger age groups.

It is important to note that there are a number of checks and balances within the modelling process that generally ensure that any inconsistencies in current or projected outcomes can be identified. In this particular case, there are at least two. First, there should be some consistency in the observed transitions for all age categories for any given country over time – which there are. If this was not the case, then a search would be carried out on the data or for country-specific events. Secondly, the proportions leaving the educational transitions modelling (e.g. age 28) should be reasonably aligned with those observed amongst 28-year-olds in the main stock-flow model – which is again the case. However, our evidence suggests that, for many countries, the qualification transitions are not quite finished at age 27 or should be truncated slightly earlier. The discussion of the modelling above indicates that there are pros and cons in extending or reducing the upper age limit, and that this would need to be considered on a country-by-country basis.

4.2.3. Combining the age- and educational transition-related results

The final piece of the jigsaw is to match the two parts of the stock-flow modelling together. This requires the data in Table 4.1 to be transposed to match the age-related stock-flows of those over 27 years of age, as shown in Table 4.3. The results of this are difficult to label, as there are changes going on in the starting years of the various cohorts, as well as the chronological years. Again, the data here are used for illustrative purposes only.

Table 4.3 Matching the proportions from the education- and age-only proportions (medium level qualifications)

		Year							
Age in 2018:	Age in 2019	2018	2019	2020	2021	2022	2023	2024	2025
									0.01
								0.01	0.04
							0.01	0.04	0.12
						0.01	0.05	0.12	0.21
					0.01	0.05	0.13	0.22	0.61
				0.01	0.05	0.13	0.22	0.62	0.72
	16		0.00	0.05	0.13	0.23	0.62	0.73	0.82
16	17	0.00	0.06	0.14	0.24	0.63	0.73	0.83	0.78
17	18	0.07	0.15	0.24	0.64	0.74	0.83	0.78	0.74
18	19	0.16	0.25	0.65	0.75	0.83	0.79	0.75	0.59
19	20	0.26	0.68	0.75	0.84	0.79	0.75	0.60	0.56
20	21	0.64	0.78	0.84	0.79	0.75	0.60	0.57	0.54
21	22	0.74	0.87	0.80	0.76	0.61	0.57	0.54	0.53
22	23	0.83	0.82	0.76	0.61	0.57	0.55	0.53	0.51
23	24	0.80	0.79	0.61	0.58	0.55	0.54	0.52	0.55
24	25	0.77	0.62	0.58	0.55	0.54	0.52	0.56	0.54
25	26	0.64	0.59	0.56	0.55	0.53	0.57	0.55	0.52
26	27	0.61	0.56	0.55	0.54	0.58	0.56	0.54	0.53
27	28	0.59	0.58	0.55	0.59	0.57	0.55	0.54	0.54
28	29	0.57	0.57	0.61	0.59	0.57	0.56	0.55	0.50
29	30	0.57	0.61	0.61	0.58	0.57	0.57	0.52	0.50

Source: The Authors.

Now, the (pseudo-) cohorts run horizontally from left to right. Taking the first column (age in 2018), this now represents the pseudo-cohort that begins at age 16 in 2018 and ages from left to right in the shaded area, which represent projected values which continue to 2035. Thus, the proportion of individuals with a medium level qualification is zero in 2018 and 0.06 (six per cent) in 2019 and 0.78 by 2025 (age 23). Note that this is not a monotonic rise, largely in line with Figure 4.1 the proportion rises to age 22 (2024) and then falls away again as individuals move from medium to high level qualifications. The individuals aged 17 in 2018 (first column) are observed from those in the previous cohort, aged 16 in 2017. In fact the proportion with medium qualifications for 16 year olds in 2017 was also zero, but can now be observed to be predicted as 0.07 in 2018. Likewise, individuals from the cohort aged 16 in 2019 (see the second column) rise from zero in 2019 to 0.72 in 2025 (before falling away about a year later).

The stepped line in the table, rising from left to right marks the dividing point between the projections of the educational-transition model and the age-related stock flow estimates shifted forward in time. It can be seen that the “join” is barely discernible in the present illustrative example and, while it can be greater in some other cases (e.g. across gender and country), it is rarely little more than the normal year on year sampling fluctuations that take place in the data. Again, reading across rows, for those 23 in 2018 (column 1) the proportion of their cohort age 16 (2011) reporting medium qualifications is 0.01, which reaches a peak of 0.86 in 2017 and, as Table 4.3 shows, falls away to 0.80 in 2018 and down to 0.52 in 2024 (note the slight rise in 2025 can also be found in *Figure 4.1*, and is explained above).

4.2.4. Applying the model to the labour force

Whilst the modelling has been discussed in the context of the population of working age, there is also the issue of the economically active population – the labour force. The same model has been applied to the distribution of qualifications across the labour force, although economic theory clearly indicates that the movement of individuals into and out of the economically active will depend on demand side factors than the distribution of qualifications across the population as a whole. In practice, differences in the results of the two models are mainly likely to stem from two principal sources: first, the propensity of individuals in the population to work or seek work will depend upon their qualification level and age; second, activity rates are subject to the level of economic activity – the availability of jobs and job opportunities.

The first of these factors is particularly important amongst younger and older individuals. Amongst younger individuals, the activity rate may be lower amongst the low qualified because they face a low demand for unskilled workers or because they feel willing and able to improve their qualification levels and therefore enter / carry on in education and training. Above a certain age – now generally much less well defined by age or perhaps country than in the past – individuals seek to retire (or die). This again may differ between qualification levels, although less clearly at the levels of qualification aggregation used here. Individuals with the highest qualifications are generally amongst the more affluent and may decide to retire earlier. As long as these propensities (e.g. to stay on longer in education or to retire earlier) are fairly systematically related to age or time, then the modelling should account for them.

The second issue is more problematic insofar as it is related to the business cycle, as opposed to long term trends, where the cycle might have more influence on one qualification level rather than another. This might particularly be the case, depending on the timing of the cycle, where, for example, low skills were adversely affected by the downturn vis a vis high skills, but benefitted more from an upturn. While the focus of the macro-modelling is longer-term than this, the EU-LFS data will reflect any affects of the cycle. Certainly using trends over a 10-year period should nullify the worst of any business cycle effects on labour supply.

In practice, the estimates and projections of the labour force appear to work reasonably well. It is possible to examine the trends in qualification levels amongst the economically active and check, for example, whether there are greater problems of projected values lying outside the 0,1 range and whether activity changes with age systematically, as describe above. However, there is a stricter test, in that the projected numbers of active individuals should always be less than or equal to the population numbers for that gender, age and country – in other words, activity rates should be less than or equal to unity and greater or equal to zero. There are some problems of this type, but they are restricted a generally fairly limited number of the low qualification groups, which were in the main already characterised by values approaching zero or amongst the high skilled, where the opposite was true. The problem was handled by imposing a restriction on the activity rate to lie between a small positive rate and below unity.

CHAPTER 5. Modelling employment growth by occupation (expansion demand)

This chapter describes the procedures to estimate the occupational and qualification employment shares. It illustrates various techniques and data sets explored to improve module 2 (EDMOD) and module 3 (QMOD) of the Cedefop modelling framework (Figure 1.1). The focus here is on changing patterns of skill demand (as measured by occupation and qualification) within sectors. Overall labour demand by sector is dealt with in the E3ME model (Chapter 3).

The chapter provides a brief review of relevant literature on occupational modelling, as well as the specifications and data used in previous research to examine how the data available can facilitate the development of a 'best practice' approach at pan- European level.

The first section focuses on relevant data and the specification of an ideal model. The second describes the l. current approach based on the much more limited data now available. The limitations of data being made available from the LFS have meant that the levels of sophistication adopted in the early stages of this work are no longer feasible. The projections are therefore based on much simpler assumptions of how occupational trends within industries will develop in the future. These build on the models and trends developed when more detailed data were available from the LFS.

5.1. Data available and the ideal specification

Most occupational forecasts at national level are based on using a macroeconomic model to project sectoral employment and then making assumptions about how patterns of employment by occupation (and sometimes also qualification) changes with these sectors change over time. Typically, these patterns are assumed to change in the future based on extrapolations from past trends, or sometimes based on expert judgement ([10]).

The occupational structure of employment is therefore projected forward taking into account the changing structure of the economy by sector and also recognising the evolution of the demand for skills as measured by occupations

([10]) For examples, results for the UK in Wilson et al (2022), which uses statistical methods to extrapolate occupational patterns within industries forward, and those for the USA which base the futures shares on expert judgement (U.S. Bureau of Labour Statistics, 2022).

within sectors. Occupational shares for each sector are calculated for the historical period, and then extrapolated forward. for the target year.

The Cedefop project adopts the trend extrapolation approach. This utilises aggregate pan-European data from the EU-LFS, as published by Eurostat. These data, covering all 27 Member States have been available since 1993, although unfortunately not on a consistent basis. Between 1993 and 2006 data were available containing information on employment by country, gender, industry (41 industries), occupation (27 occupations, based on ISCO88) and qualification (3 level qualifications). These data measure the changing pattern of skill demand ([11]) and are in theory harmonised to a common classification.

These data were used to create a consistent data set of employment by sector, occupation and for all countries. This data set was constrained to match the sectoral data from E3ME (based on National Accounts) while at the same time mirroring the occupational and qualification employment patterns exhibited in the LFS.

Using these data simple trend extrapolative methods were developed for projecting occupational employment patterns within industries. Following extensive analysis, the preferred specification was simple logistic time trend, fitted on published LFS data for the period 1993-2006:

$$Ln\left(\frac{S}{1-S}\right) = a + b * Time$$

Despite considerable effort to develop more general models, the most robust results were obtained using simple country data sets, distinguishing occupation and qualification for each sector, based on the published EU-LFS shares and using models that included just ‘time’ as the only independent variable.

Where data are unavailable or inappropriate (due to missing or inconsistently classified information resulting from changes in classification or other discontinuities), the estimation period is truncated accordingly. An algorithm also checks to see if the projected changes are plausible, censoring out shares that lie outside the range zero to unity and also where the projected change is exceptionally rapid (which usually arises as a result of idiosyncrasies in the data). Where there were problems of this kind, the algorithm explores alternative specifications (log linear, linear and fixed shares) until an acceptable outcome was achieved. This applies to both shares of occupations within each sector and shares of qualifications within occupations.

([11]) However, the observed employment levels are the consequence of both demand and supply factors. The latter may be especially important with regard to the qualification dimension.

5.2. More recent developments in modelling

Unfortunately, it has not been possible to extend the LFS dataset on a consistent basis. In 2008 a new system for classifying occupations was introduced (ISCO08). No general crosswalk between the ISCO88 and ISCO08 was available for all countries.

This introduces a major discontinuity int the trends in skill patterns within sectors. To deal with this it was assumed that future trends in skill patterns within sectors based on ISCO08 would follow those previously established for ISCO88 occupations, but applied to the new shares of employment as measured using ISCO08.

As new information becomes available on the new system of classifying occupations it should in principle be possible to begin to discern the latest trends. Currently, there are now 12 years of data available on the more recent ISCO08 system of classification. However, concerns about confidentiality have led a number of countries to restrict release of detailed EU-LFS data. This means that it is no longer possible to identify changes in occupational employment patterns within sectors at the same level of detail as was possible in the original modelling work summarised in Cedefop (2012).

The current set of results are therefore based on a much more broad-brush approach. This used the same trends within sectors (as established in the earlier modelling work) but constrained to match the overall trends in occupational employment ***across all sectors*** as shown in the LFS.

CHAPTER 6. Modelling replacement demand

In addition to analysing changes in overall occupational employment levels, it is important to consider replacement needs arising from retirements, net migration, movement into other occupations and in-service mortality. This is referred to as replacement demand. In general terms, replacement demand can be seen as job openings arising because of people leaving the workforce or their occupation. Most work on replacement demand has tended to focus on 'permanent or semi-permanent' withdrawals from the employed workforce. The main reasons for this are retirement, emigration, and especially for women, family formation and child-bearing and rearing. Next to withdrawals from the labour force, we also include inter-occupational mobility – movements of workers from one occupation to another. These movements leave a vacancy and hence create a demand for a worker to fill the vacancy.

The first section of this chapter provides the general overview of the replacement demand concept. The second section is devoted to cohort component methods ([12]) used to produce the results since the Methodological report (Cedefop, 2012). The current specification of the model is described in the last section.

6.1. General overview

The main data used to model replacement demand are based on the EU-LFS. These data allows us to analyse the demographic composition of each occupation. It allows us to estimate specific rates of outflows for each occupational class. EU-LFS data can also be used for making estimates of outflow rates.

The replacement demand model (RDMOD from Figure 1.1) has been developed based on similar data sources to the occupational model (ECMOD). It is driven in part by the demographics (age-gender composition) of employment in combination with the occupation (and age-gender) specific outflow probability due to retirements, occupational mobility and migration. The model used builds upon the national model from the Netherlands (Cörvers et al., 2008). Similar models using variants of the methodology are employed in several other countries, both within and outside of the EU.

Projections of occupational employment typically focus on the total numbers of people that are expected to be employed in such jobs in the future. While such estimates can provide a useful indication of areas of change, highlighting the likely

([12]) Technical details about this method are provided in Kriechel and Sauermann (2010).

net ‘gainers’ and ‘losers’, they give a misleading impression of job opportunities and skill requirements. Even where the projections indicate significant employment decline over the medium term, there may be good career prospects with significant numbers of new job openings. This is because, as long as significant numbers of individuals are still likely to be employed in the future, employers will need to replace those employees who leave for retirement, career moves, mortality, or other reasons. Replacement demand may often dwarf any ‘structural demand’ or ‘expansion demand’ resulting from growth in employment in a particular category. It can easily outweigh any negative changes due to the projected employment decline.

While the concept of replacement demand is simple enough to grasp, estimating it is a different matter. The main problem is that official statistics emphasise measuring stocks of people in particular states rather than flows, which is essential to estimating replacement demand.

However, use can be made of readily available statistics to provide indicative estimates. Ideally, one requires a full set of demographic accounts that trace people’s movement from one socioeconomic position (e.g. employment in a particular occupation) to another (e.g. retirement). In practice, such a complete set of accounts are rare, even at the national level. However, for several consecutive years the LFS has provided a sufficiently large sample to obtain estimates of the main elements at the national level. The key components are: information on the age and gender structure of occupational employment; information on rates of outflows due to retirement (including early retirement); inter-occupational mobility; migration and/or other reasons for leaving the workforce.

The information on outflow rates can also be estimated using stocks of age-cohorts within occupations for several years. Using the year-to-year changes the outflow rates by occupation-age cohort can be estimated. However, these estimates may not allow for discrimination between the reasons for the outflow that leads to replacement demand. Given the data availability, this methodology is used for the purposes of the current forecast.

Availability of following data is considered as useful for the purposes of this methodology:

(a) Age structure.

Data on age structure are required since many of the flows, especially retirements, mortality and occupational mobility, are age specific. Age structures vary significantly by occupation. For some groups, such as corporate managers and administrators, experience is a key requirement and this is associated with age. The proportion in the 45-59 year old category is therefore relatively high. In contrast, in many other occupations the age structures are much more heavily skewed towards younger age groups. In sales occupations, for example, the age structure is much more heavily weighted towards younger age groups. Differences in age structure across occupations influence replacement demand due to occupational mobility and retirement, which are age related. Even inter-occupational mobility is affected differently over occupations;

(b) Retirement.

Retirement rates vary by gender and by age and may differ for different occupational groups. But since sample numbers are often too small to allow for meaningful estimates methods to deal with these problems need to be adopted. Estimates can be based on data from the LFS, which show the percentage of those employed one year ago who have retired from employment, either temporarily or permanently. For males, the main outflows are associated with retirement per se. For females, in particular, there is significant, quite often temporary, outflow for younger age groups associated with family formation;

(c) Mortality.

Another potential outflow is due to mortality. While losses due to death are not great for individual age groups up to the age of 65, they can cumulate to produce significant losses over an extended period. However, the current model does not explicitly incorporate differential mortality risks (not least because no significant or radical changes are expected in them). Rather the focus of the cohort component methodology is to identify overall outflows over cohorts, irrespective of the cause (sickness, death, family obligations);

(d) Migration.

Net migration can have an important effect on the in- or outflow of the labour market. One of the problems of migration is the lack of suitable data;

(e) Occupational mobility.

Occupational mobility is an important source of replacement demand in some occupations although not for all. The full occupational mobility flow matrix indicates that some occupations such as managers tend to gain employment as people are promoted from other occupations. The cohort component approach does not differentiate the replacement demand which is due to occupational mobility. It only identifies net mobility.

The overall scale of change is obviously dependent upon the length of period considered, as well as the opening stocks and the age structure of the current workforce. Replacement demand is also dependent on the level of occupational aggregation. With lower levels of aggregation, the observed occupational mobility is lower. For most projections rates of outflow are assumed to be constant over time. The scale of structural or expansion demand (which in some cases may be negative) is usually modest compared to replacement needs, and in most cases the latter offsets any negative change.

Replacement demand is driven by the proportion of employees in an occupation that is likely to leave within the forecasting period. A higher share of those workers in the base period of a forecast will lead to a higher predicted replacement demand. Business cycles are also prone to shifts and outflows at some point in time. With respect to the timing of (early) retirement, in good economic times people are likely to stay on a little longer; hence the outflow rate will be somewhat lower. In contrast, in bad economic times, people that are likely to leave soon will have incentives to move out earlier.

The period of the economic downturn (global or sectoral) has an impact on the timing of the outflow that can lead to some short-term shifts in replacement demand. Companies try to accommodate the lower demand for workers by reducing flexible work (temporary agents and not core stuff), but also by bringing forward outflows that are likely to occur in the near future (retiring rather than firing). In other words, early retirement schemes – official or not – are being used to reduce the workforce if possible. This implies that the outflow among the cohort of older workers, who are close to retirement, is temporarily higher, while the outflow will be lower for some time after the crisis. This lowering effect is the simple result of the reduction in the population of workers that reach retirement age.

While the outflow increases temporarily – this in the methodology of replacement demand means that the replacement demand increases as well – it is not likely that replacement demand in those economic circumstances will be filled in immediately. Some catch-up will take place after the economic recovery. The annual replacement rate is taken as an average over the entire forecasting period, which is not affected by such cyclical behaviour on the labour market. However, it is important to be aware of the implications of the crisis for in- and outflows of workers that might lead to a temporary deviation from the overall replacement rate.

6.2. Current specifications of the model

The methodology is based on the cohort-component analysis that uses the EU-LFS for all countries, while disaggregating education into several ISCED categories (for replacement demand by education) and ISCO categories (allowing estimates of replacement demand by occupation).

There are three components to the model:

(a) a forecast of demographic development within a country;

(b) a forecast of (changes in) participation, preferably by gender and age groups;

(c) an estimate of the outflow by occupation (education) category, gender and age group.

Components (a) and (b) are usually considered external to the replacement demand model. The most recent demographic forecast by Eurostat are used to produce the estimated. Changes in participation use the same participation rate by country, age and gender as generated within the E3ME model. This insures consistency across the entire set of forecasts.

The basic steps use occupation (subindex o) as the relevant subcategory. However, one can interpret the methodology analogously if education is used instead. For the purposes of the Cedefop methodology, education is not estimated separately given the high level of aggregation on the education variable. Rather, the replacement demand by education is deduced from the occupational replacement demand. By using the occupational replacement demand and imposing the most recent distribution of education by occupation, we are able to present the most likely replacement demand using the current demand for education levels within an occupational class. A similar approach has been applied in determining sector-specific replacement demands.

Table 6.1 gives a schematic input-output table of the labour force/population in a country (see also Willems and de Grip, 1993). The first rectangle gives the movements within the labour market. The second, bigger rectangle encompasses movements out of the labour market, while the third rectangle also considers changes in the population. Adding rows (for time t) or columns (for time t-n) of these flows gives the total population within an occupation.

Table 6.1 Schematic of replacement demand

Outflows **Inflows**	**Occu-pation 1**	**Occu-pation 2**	**Unem-ployed**	**Outside the labour force**	**Outflow populations**	**Total**
Occupation 1	A	B	C	D		**W1,t-n**
Occupation 2	E					**W1,t-n**
Unemployed	F					
Outside the	G					

labour force						
Inflow population						
Total	**W1,t**	**W1,t**				

Source: The Authors, based on Willems and de Grip, 1993.

Several flows are indicated in the table with capital letters. A denotes the workers that work in occupation 1 at time t-n and continue to do so in period t. B denote the workers that move from occupation 1 to occupation 2 in the observed time. E denotes the opposite movement from 2 to 1. Thus, B and E denote the job-to-job mobility. C and D denote movements out of the labour market from holders of occupation 1. Corresponding inflows into occupation 1 are F and G in the schema.

The first step in modelling future replacement demand per occupational class is a description of the inflow and outflow patterns by occupational class in a historical period. Because there are no appropriate data for mobility flows on the labour market, stock data are used. With the cohort components method, cohort-change rates based on the number of persons of the same birth cohort who were employed at two different time periods can be calculated (Shryock and Siegel, 1980). These cohort-change rates can be rewritten as average annual net inflow or outflow percentages (flow rates for males and females are differentiated):

$$\dot{F}_{o,a}^{t-1} = \frac{k\boldsymbol{W}_{o,a+1}^{t} - k\boldsymbol{W}_{O,a}^{t-1}}{k\boldsymbol{W}_{o-a}^{t-1}} \qquad (1)$$

Where $\dot{F}_{o,a}^{t-1}$ is the annual net inflow or outflow ratio of workers in occupational class o of age group a (with class width k) at time t-1 during the period (t-1,t); $\boldsymbol{W}_{o,a}^{t}$ is the number of people working in occupational class o of age group a (with class width k) at time t. The time lags are used to describe how particular cohort is moving within an occupation. If $\dot{F}_{o,a}^{t-1} > 0$, there is a net inflow for a certain age group from an occupational class, and if $\dot{F}_{o,a}^{t-1} < 0$ there is a net outflow.

The second step in modelling is to translate these inflow- and outflow-percentages into the replacement demand by occupational class. For occupational classes with an increase in employment in the period (t-1,t), replacement demand is equal to total net outflow in this period. However, for occupational classes which faced a decrease in employment, not all vacancies created by the outflow of workers will have been filled by new workers. Therefore, replacement demand for these occupational classes equals the number of vacancies likely to be actually refilled, which is to say the total inflow of workers into the occupational class. In this way, the more or less ‘structural’ replacement demand is derived. This methodology measures only the net flow to or from an occupational class. This means that replacement demand satisfied by re-entering workers of the same age cohort is not measured. So replacement demand is actually measured only for newcomers to the labour market.

A model is then estimated in which the net inflow or outflow ratios are explained on the basis of the average inflow or outflow from the total working population on the one hand, and the occupation-specific deviations per age-gender group on the other. This approach guarantees that the sum of the net flows among the occupations corresponds to the total inflow or outflow. Written mathematically:

$$\dot{F_o} = \dot{F} + \sum_k \beta_{ok} D_x \qquad (2)$$

Where: $\dot{F_o}$ is the vector of net inflow of outflow ratios for occupation o, with observations for gender, age group and year; $\dot{F}$ is the vector of net inflow – outflow ratios for the total working population; D_x is the matrix with dummy variables where elements are equal to 1 for cohort x and 0 elsewhere; β_{ok} and represents random parameters.

Finally, a projection is made based on the estimated coefficient combined with participation rate and population changes applied to the age-gender structure of the occupation as predicted by demographic and participation forecasts. To model the demographic composition of an occupation and its dynamic changes, we project uniform changes of one age cohort in the next cohort over the full time-horizon of the forecasts. We then get the following model:

$$\boldsymbol{W}_{o,a}^{t} = \boldsymbol{W}_{o,x}^{t-1} \, (1 + \boldsymbol{P\dot{O}P}_{o,x}^{t-1})(1 + \boldsymbol{P\dot{A}RT}_{o,x}^{t-1}) \qquad (3)$$

Where $\boldsymbol{P\dot{O}P}_{o,x}^{t-1}$ ($\boldsymbol{P\dot{A}RT}_{o,x}^{t-1}$) are the annual net changes in population (participation) of workers in occupational class o of cohort x at time t-1during the period (t-1,t).

The replacement demand is derived from a combination of the estimated outflow coefficients (based on historical outflow between 2010-2014) per age-gender-occupation and the predicted change in participation rates per age-gender. In a simplified manner, RD is calculated as follows:

$$RD_{x,o} = -NET\ FLOW_{x,o} \boldsymbol{W}_{x,o}^{t} \qquad (4)$$

The estimated net flow is entering RD with a negative sign as a net outflow implies that more replacement is needed. The estimated net flow is then multiplied with the expected number of workers in time t.

These results generate annual rates of replacement demand outflows by occupation and broad level of qualification for each country over the projection period. These rates are assumed fixed over all years and are applied across the board for all sectors.

In reality, we would expect these rates to vary depending upon the actual demographics in each sector. For example, all else equal, those sectors with a greater proportion of older workers nearer to retirement age would be expected to see higher replacement demands than those with a younger workforce.

Based on expert judgement, adjustments to the benchmark projections can be made, i.e. higher or lower rates of replacement needs in different occupations. The important message is that it is not just expansion demands that are important in assessing future skill needs. Replacement demands generally outweigh expansion changes by an order of magnitude.

CHAPTER 7. Reconciling skills supply and demand

Economic theory as well as economic reality shows that coexistence of unemployment and unfilled vacancies is a natural and simultaneous feature of the labour market. Even in full employment it is impossible to reduce the number of unfilled vacancies to zero due to the search costs and asymmetries in the labour market. This is especially the case when the imbalances within particular occupations are caused by lack of appropriate skilled labour force, which is usually solved by retraining and/or reallocating workers. One of the aims of the process of anticipating changing skill demand and supply is to uncover potential imbalances in such a way that adequate preventive measures can be taken.

Knowing more about potential skill imbalances in the labour market is very important for policy. The projections of the skills supply and demand can throw light on possible future developments in European labour markets, highlighting potential education mismatches and helping to inform decisions on investment in skills, especially formal qualifications, made by individuals, organisations and policy-makers.

Comparing demand and supply projections is problematic for both practical and theoretical reasons. Unless the two sets of results are based on common data and are carried out simultaneously, they cannot be directly compared. Many adjustment mechanisms operate in the labour market to reconcile any imbalances that may arise. In the short term, these include adjustments in wages and different kinds of mobility, as well as changes in the ways employers use the skills that are available. In the longer term both supply and demand will adjust to reflect the signals and incentives that arise due to these forces.

Generally, employers will not cease their operations if they cannot find the ideal mix of skills. They will operate with what is available. Conversely, if the educational system delivers too many people with particular levels of formal qualification this does not necessarily imply that these people will remain unemployed. Rather they tend to find jobs that make less direct use of their qualifications, since their education and training still often puts them at an advantage in the labour market compared to those with lower level qualifications. The labour market operates as a kind of sorting mechanism that allocates people to jobs, based on the limited information available to both sides (employers as well as actual and potential employees).

In principle, these mechanisms can be modelled and incorporated into a forecasting tool. In practice, this demands very detailed and rich data that are, at present, not available at pan-European level. In the present results, a more limited reconciliation is used to recognise the key features of the labour market interaction. The projected numbers on the supply side are taken as given for this purpose. This reflects the fact that the total numbers available by qualification level are largely predetermined by demography and educational and training decisions already made. The much better qualified new entrants coming into the labour market, replacing much less well qualified older people, imply that substantial improvements in average qualification levels are inevitable in the short to medium term.

This chapter describes Cedefop solutions to the above-mentioned problem. The main data and methodological limitations of reconciling labour supply and demand with respect to qualifications, including the use of labour market accounts residuals are presented in Section 7.1. Section 7.2 focuses on the approach used by Cedefop to provide reconciliation of skills supply and demand. The construction of possible indicators for better understanding and interpretation of the imbalances is described in Section 7.3.

7.1. Measuring imbalances

The current Cedefop framework focuses on employment in sectors, distinguishing occupation and highest qualification held; labour supply by qualification, age and gender; and unemployment distinguished by qualification category.

When comparing estimates of demand and supply of labour (and skills) it is important to recognise significant problems that arise due to different measurements applied in different sources. This section focuses on differences in the historical estimates of the labour force, employment and unemployment used in the Cedefop framework.

The population data used in E3ME are constrained to match the official Eurostat numbers and projections. The labour supply numbers in E3ME are therefore not LFS benchmarked as such, although they use LFS activity rates. They rely on Eurostat demographic data to produce the overall numbers and LFS activity rates by age and gender.

Labour demand is measured as employment. Obviously this is not strictly correct. Observed employment is the consequence of both demand and supply factors, nevertheless it is common practice in work of this kind to refer to employment as de facto ‘demand’. Overall employment levels from E3ME by sector are translated into implications for occupations by the EDMOD module, as described in Chapter 5. This is based on a detailed analysis of occupational employment patterns within sectors. This module also produces initial (unconstrained) estimates of employment by qualification (again based on an analysis of changes in employment patterns within occupation and sectors).

To focus on imbalances by skill (as measured by qualification), these initial estimates of employment by qualifications are constrained to match 'supply in employment'. This is a measure of the number of people by highest qualification held who are economically active and in employment. This is on an EU-LFS demographic accounts basis. It is then scaled to match the E3ME NA based estimate of employment. This final set of employment estimates by qualification matches the total employment in E3ME on a NA basis and is known as 'constrained demand'.

The LMAR in E3ME is the difference between employment (workplace jobs, NA) E(j), and employment (head counts, LFS based, residents) E(r).

$$LMAR = E(j) - E(r) \quad (5)$$

The person/jobs distinction is not entirely clear cut. The LMAR arises for a range of reasons of which occupation is just one (double jobbing). Other factors include: commuting (flows across national borders) and statistical discrepancies between NA and LFS based estimates. NA employment appears to generally refer to persons rather than jobs (as in the LFS), meaning that most of the discrepancies arise from reasons other than double jobbing.

The LFS based estimate of employment (E(r)) is a count of people employed in a particular country. It is equal to the labour force (LF) less the number of people unemployed (U), both measured by head count.

$$E(r) = LF - U \quad (6)$$

Information on unemployment by qualification level is available from the LFS. This can be used in combination with the projected totals of unemployment from E3ME to generate projected levels of unemployment by qualification level, by making assumptions about maintenance of historical differentials in unemployment rates. Equations 5 and 6 can be variously rearranged to derive indicators of interest from the model outputs:

$$E(j) = LMAR + (LF - U) \quad (7)$$

In E3ME the following slightly different terminology is used:

$$E(j) = EMP \quad (8)$$

'Unemployed' in E3ME is given by:

$$Unemployed = LF - E(j) = E(r) + U - (LMAR + E(r) = U - LMAR \quad (9)$$

E3ME then generates a variable called ‘U level’ as follows:

$$U\ level = Unemployed + LMAR \qquad (10)$$

This is equivalent to the normal ILO measure of unemployment.

Unemployment is calculated this way in all countries, except for Luxembourg where the discrepancies were so large that it was not possible to model unemployment in a stable manner. The LMAR is calculated to match the last year of historical data, taken from AMECO database of European Commission’s Directorate-General for Economic and Financial Affairs, which uses ILO-consistent definitions. In the modelling, the LMAR is calculated using a ‘fixed share of employment’, although in the workbooks it can appear to vary. This is because the unemployment rate in the workbooks is only calculated to one decimal place. ‘U level’ can easily be calculated from ‘U rate’ by multiplying by the total labour force (LF).

However, as shown in equation 10, this will include the LMAR. The value of the LMAR will vary for those with high, medium and low qualifications. It is possible to assess this by going back to the raw LFS data and using the historical estimate of the unemployment rate for high-, medium- and low-qualified people to generate a number for those unemployed in each category in the base year. This can then be compared with the difference between the demand and supply numbers in the base year. The difference is the LMAR for that qualification category. In principle, this can then be projected assuming (in the first instance) that it remains a constant proportion of the total LMAR (from E3ME) for all future years. In practice this has not been done. Instead, an assumption is made about how unemployment is ‘shared out’ among high, medium and low qualification categories.

The E3ME data on total unemployment are used to provide overall constraints on the implied unemployment numbers by qualification (three broad levels: high (H), medium or intermediate (M), and low (or no) qualifications (L)) in the imbalances work. The latter are constrained to match the overall E3ME estimates of ‘U level’, making assumptions about how overall unemployment will be ‘shared’ between those with different levels of qualification, using LFS historical data and assumptions about how these patterns might change in the future. These assumptions are that previous differentials will be maintained but that the better qualified will take an increasing share of total unemployment in line with their increasing share of the workforce.

Given these levels of unemployment by qualification level, it is possible to generate constrained estimates on those in employment by qualification level.

These can then be compared with the unconstrained estimates obtained by extrapolating trends in employment shares.

A measure of initial imbalance and mismatch by H, M and L qualification levels can then be obtained by subtracting the unconstrained 'demand' estimates (employment estimates by qualification from the demand model, summed across all sectors and occupations) from the corresponding 'supply' estimates (the total labour force across all ages and genders).

Initial comparisons suggest that the independent projections of qualifications patterns for supply and demand exhibit significantly different trends . The supply of those with high and intermediate level qualifications generally rises more rapidly than the demand.

The implied demand side employment rates by qualification category, when compared with the supply side activity rates, can sometimes indicate odd patterns across categories and over time. These reflect inadequacies in the basic data. In principle, the employment rate should be less than the activity rate (the numerator of the latter being employment plus unemployment). In such cases, a final adjustment has been made to the employment by qualification estimates to take into account the LMAR.

7.2. Estimation of imbalances

The historical patterns of employment by qualification observed are the result of a combination of both supply and demand factors. Separating them is not straightforward. Recent trends have seen a sharp rise in the formal qualifications held by those in employment in most countries. There is some evidence that this reflects demand changes, with many jobs requiring more formal higher level qualifications than before. There are also indications that the returns on such qualifications have remained high (for a review, see Wilson et al., 2007). On the other hand, it is clear that there have been major changes on the supply side, in part at least in response to government policies to increase participation in higher education. The latter has greatly increased the number of people entering the labour market with high formal qualifications. The proportion of young people with higher formal qualifications is much higher than for older people. There is, therefore, a strong cohort effect. This has been reinforced to some extent by increasing qualification rates for older people (due to lifelong learning and upskilling effect) (part of Section 4.2).

The overall supply of people holding formal qualifications at higher level (ISCED 5-8) is relatively straightforward to conceptualise and model. However, there are considerable conceptual and practical difficulties in extending this to include lower-level qualifications (ISCED 1-2). A particular difficulty is to conceptualise the idea of supply to cover specific dimensions such as occupation, sector and geographical area. This is because the educational systems in most countries are not completely hierarchical. These issues are discussed in turn:

(a) limitations by ISCED level:

The first problem to be addressed in extending this type of model to cover the full range of qualifications is the much more limited information available on lower level qualifications (ISCED1-2). Ideally, stock-flow modelling requires a comprehensive set of demographic accounts showing how individuals progress throughout the educational system and the labour market over time. In practice, such accounts do not exist, although there is a considerable amount of information on certain flows as exploited in the skills supply module described in Chapter 4.

Statistical agencies collect and publish a considerable amount of information on the higher education in particular. This can be used to develop estimates of the main flows involved and thereby develop stock-flow models of this process. In practice, this information is often disparate and far from comprehensive.

In the case of lower-level qualifications, while there is an enormous amount of detailed information available on the acquisition of qualifications, there is much less information on what prior qualifications these individuals may have possessed. This makes it difficult, if not impossible, to develop stock-flow models analogous to those constructed for higher levels (e.g. for the UK in Wilson and Bosworth, 2006).

(b) highest versus all qualifications held:

The discussion so far has focused on highest qualifications held. As individuals acquire ISCED 4 and above qualifications, it is almost inevitable that the proportions with ISCED 1-3 as their highest qualification will fall. This can mean that, despite increases in those acquiring ISCED 1-3 qualifications, the numbers and proportions of people possessing these as their highest qualification may actually decline.

(c) problems in conceptualising supply into occupations or sectors:

Most jobs (and thus occupations) are undertaken by people with a range of formal qualifications. This is partly a function of age, with older workers generally relying more on experience than formal qualifications. However, even allowing for the age factor, there are enormous differences. This makes defining the supply of people into an occupation almost impossible. It is possible to identify some key elements, focusing on the flows of people through the education and training system, but boundaries are too blurred and transitory to enable robust quantitative modelling.

Much the same is true for the concept of the supply of labour to a sector. This will depend upon the occupational mix of the sector and its geographical location. For some occupations, the labour market may be worldwide. This is increasingly true of many high level managerial and professional groups. Ever increasing ease of transport now means that it is also a feature of the labour markets for many lower-level occupations (for example, construction and agricultural workers, as well as nurses). While these issues may be addressed within individual sectors, it is very difficult to develop a general approach that can cover all these aspects consistently for the whole economy.

There may be some overqualification or underqualification, this pressure varies among occupations and sectors. The following elements are important here:

(a) the demand for qualifications model, which delivers overall numbers of people in employment, qualified at three broad ISCED levels;

(b) the stock of labour supply (numbers of people actively searching for work);

(c) the sorting model (SORT algorithm described in Annex 3), which sorts people according to the qualifications held into occupations to make the results from (a) and (b) consistent.

The SORT algorithm reconciles the two sets of estimates of demand for and supply of qualifications. This final element compares the supply numbers with the demand ones and re-computes the employment patterns to bring the two into agreement (making certain assumptions about unemployment). Effectively it acts as a sorting mechanism that raises or lowers qualification shares within occupations until demand and supply numbers match. This does not imply that demand and supply are in balance however, since some people may be overqualified or underqualified for the jobs they are employed to do.

The SORT algorithm used in BALMOD ([13]) provides a simple approach to reconciling aggregate demand and supply results, given the data available for all countries. Its outcome is that for individual occupations or sectors the patterns of qualifications as revealed by the original unconstrained demand projections and the constrained ones will show how any surpluses or shortages affect the qualification mix. If supply is growing faster than demand for particular levels of qualifications, the constrained qualification mix will be ‘richer’ than the

([13]) Module BALMOD of the Cedefop conceptual modelling framework (Figure 1.1) is designed to bring together the demand and supply estimates to tackle this issue. Demand is projected by sector, occupation and qualification level (highest qualification held), whereas the supply focuses on qualification by gender and age.

unconstrained ([14]) one (and vice versa). Comparison of the constrained and unconstrained results provides a useful indicator of supply-demand pressures for different occupational and sectoral groups.

The concept of the supply of qualifications at spatial level is somewhat more manageable than it is for occupations or sectors. It is relatively straightforward to develop quantitative estimates and projections of population and the labour force for each country. In principle, this can be extended to cover formal qualifications held. However, the data available at pan-European level are generally less robust than at national level. Moreover, the issues of commuting and migration flows become significant. Therefore, the modelling is limited to a more simplified level than the more detailed and sophisticated stock-flow analysis applied in some individual Member States (e.g. the UK, Wilson and Bosworth, 2006).

7.3. Indicators of imbalances

In addition to the reconciled demand and supply estimates, indicators of imbalances were also developed to help to understand how imbalances evolve, to identify emerging trends and to allow for comparison among countries. These indexes provide a comprehensive picture of possible imbalances and mismatches in the labour market. First the interpretation of unemployment as an indicator of mismatch is discussed. Then we discuss two indicators that are used in the country fiches. The third subsection discusses the remaining indicators which are presented in the workbooks. In a final subsection we discuss some general thoughts on the use of imbalance indicators.

7.3.1. Unemployment as a general indicator

The first set of indicators is based on the supply and demand forecast from the separate modules. These indicators are indicative of overall imbalances in supply and demand, i.e. an oversupply or undersupply, be it by education or in total. Furthermore, we use the extrapolated unemployment rates by education level, both as an indicator of the allocative process but also of imbalances that are not directly identified within the model.

The main relation of labour supply and demand is that supply should be equal to the sum of fulfilled demand and unemployment. As described in previous sections, unemployment is derived within the model. The unemployment rate by

([14]) As unconstrained demand we understand demand initially produced by the model. As constrained demand we understand demand adjusted to fit the supply.

education level can be seen as an indicator of matching efficiency. Overall high levels of unemployment imply that the supply cannot be properly matched to demand. This can have two reasons: (a) there is too much supply relative to demand; (b) the supply does not meet the requirements of the demand. The first point is straightforward as it a simple imbalance of total oversupply. The second one captures elements of the efficiency of the labour market in matching and generating the supply of skills that the economy actually needs. For example, there can be a high overall unemployment rate if the labour market demands only highly educated workers, while the supply consists mainly of people with low and intermediate levels of education. Suppose firms do not adjust their production technology to substitute low- and intermediate-educated individuals for highly educated ones, in other words, change their demand. In that case, it will lead to unemployment among the low and intermediate educated, while shortages of highly educated workers appear. Within education levels, there can also be a mismatch if the field of education within a level does not match supply and demand. For example, if the supply is mainly within the field of health, while the demand is mainly in the field of technical studies, there will likely be at least temporary unemployment. This last aspect of field mismatch is not currently modelled in the supply and demand forecasts. If, however, this is a structural problem within the economy, this will show up in higher unemployment (and vacancy) rates.

7.3.2. Indicators in the Country Fiches

The two indicators used in the final figure of the country fiches are derived from ideas of the existing imbalance indicators. The indicator of future hiring difficulties is a derivation of the IFIOD. Yet, the direction is turned around to ease the interpretation of the diagram (increasing difficulties or problems along the axis). The degree of hiring was inspired by the measure of change.

Indicator of future hiring difficulties

This indicator should identify the difficulties an organisation will likely have if it needs to hire a worker for a specific occupation. This measure indicates increasing difficulties in fulfilling demand given the available supply of qualifications used in the occupation. It is based on the unconstrained demand by qualification which is evaluated to the supply by qualification. The indicator thus summarises the supply-demand relationships of all education types that can be employed for each level of education. This is weighted by the likelihood that an occupation is filled with a certain type of education. The weighting is based on observed (base year) shares of the occupation-education matrix. The number of people working with a specific background (education) in an occupation thus determine these weights.

$$IFHD_i = 1 - IFIOD_i \quad (11)$$

$$p_i = min\,(1, \frac{supply_i}{demand_i}) \quad (12)$$

$$IFIOD_j = \frac{\sum_i p_i x_{ij,base}}{\sum_i x_{ij,\text{base}}} \quad (13)$$

Where x_{ij} is the total amount of people in occupation j with education type i in the base year. The indicator gives the relative degree to which difficulties can be expected in meeting the occupation in demand. Note that the share is the same for all occupations, as it simply denotes the relative demand to supply of an education type.

The implicit assumption is that all shortages in education types will be felt in the same way by all occupations but weighted to the importance of that education type for the respective occupation. A value of 0 indicates that there are no shortages expected, whereas a (theoretical) 1 would indicate that no demand could be fulfilled.

Degree of hiring

The degree of hiring required in the occupation measures the distance between the current composition and level of employment in a specific occupation towards the base year levels. The indicator measure for each qualification level the amount of (net) hiring needed to reach the forecasted level of employment. These changes (degree of hiring required) can be due to a change in the qualifications required or increases in the number employed.

The indicator is calculated for each occupation j across all education types i.

$$DOH = \frac{\sum max\,(0; xij,t - xij,t{-}1)}{\sum xij,t{-}1} \quad (14)$$

The indicator shows the level of change that is necessary relative to the path that the occupation is taking in the projection years. Higher values indicate a higher level of constraint and more adjustment to the current path of employment. For each occupation j and education type, the base year employment x_{t-1} is compared to the employment level at the end of the forecast: x_t. Any positive change is taken and summed over all education levels. The total of those positive changes is evaluated to the total of base-year employment across education level.

7.3.3. Indicators in the Country Workbooks

The following indicators are shown in the country workbooks. They have been developed in the past rounds of the forecasting project. Several are considered as experimental, yet we would like to provide a complete definition of all indicators and their interpretation.

Measure of change (MC)

The measure of change calculates the distance between constrained demand (D^c) to the base year counts (D^{base}). The changes are summed up by qualification level i and occupation j, and the indicator is calculated for each occupation j.

$$MC_j = \frac{\sum_i (D_{ij}^c - D_{ij}^{base})}{\sum_i (D_{ij}^{base})}$$

Higher levels indicate higher level of adjustment relative to the current (base year) state of the labour market. The indicator thus provides an insight into the size of the changes at the occupation level relative to the overall employment in an occupation.

Change in Average Qualification (CAQ)

The Change in Average Qualification (CAQ) is based on the measure of change. It is based on evaluating the average qualification level within an occupation at two periods of time, the base period and the final period of the forecast. Average qualification *i* is in our case based on the three basic ISCED levels: low (with the value 1), intermediate (with the value of 2), and high (with the value of 3).

$$CAQ_j = AEL_j^c - AEL_j^{base}$$

Where the AEL is the average education level of a specific occupation. Given the distinction in level that we can make, the AEL can take on values between 1 and 3.

Indicator of constraint (IC)

The indicator of constraint provides an insight into the differences between the evolution of the unconstrained demand model and the outcome of solving supply and demand in the constrained outcome. It thus provides the degree to which the initial demand deviates in terms of qualification mix from the feasible outcome. Technically it gives the percentage of change (relative to the constrained demand) that needs to be adjusted in absolute terms to reach the level of constrained demand (D^c) from the unconstrained demand (D^u). which is the result of the adjustment procedure. High levels of the indicator of change indicate significant adjustment processes necessary. The indicator is calculated for each occupation *j* across all qualification levels *i*.

$$IC_j = \frac{\sum_i (D_{ij}^c - D_{ij}^u)}{\sum_i (D_{ij}^c)}$$

The indicator shows the level of change that is necessary relative to the forecasted path that the occupation is taking in the projection years in order for imbalances to be resolved. Higher levels indicate higher levels of constraint, and more adjustment to the current path of employment.

Indicator of Constrained demand by Qualification (ICQ)

The Indicator of constrained demand by qualification level (ICQ) measures in how far constraints within occupations are concentrated in specific qualification levels. To be precise, the average level of constraint by qualification is calculated. Values of or close to 1 indicate that the constraint is predominantly at the lowest level, and values close to 3 indicate that the constraint is predominantly at the highest qualification level.

$$ICQ_j = \frac{\sum_i \left(\max(0; D_{ij}^u - D_{ij}^c) \cdot i\right)}{\sum_i \max\left(0; D_{ij}^u - D_{ij}^c\right)}$$

With i taking on the values of 1, 2, or 3 for, respectively, the lowest, intermediate, and highest level of qualification.

Indicator of Overeducation (IOE)

The indicator is intended to evaluate the issue of overeducation. The indicator evaluates the constrained demand (*D_c*) with respect to the qualification level. Higher values of the indicator indicate that initial demand by qualification, unconstrained demand (*D_u*), is filled with higher qualified supply. Negative values indicate undereducation or skill shortages. The indicator thus measures in how far higher qualified workers are forced to substitute for intermediate or lower qualified workers, which were initially demanded.

$$IOE_j = \sum_i \frac{(D_{ij}^c - D_{ij}^u) \cdot i}{D_{ij}^c}$$

Indicator of Future Imbalances of Demand (IFIOD)

This indicator is calculated for each two-digit occupation and denotes the difficulties an organisation is likely to confront in hiring a worker for a specific occupation. The indicator summarises the overall supply-demand relationship of qualification levels weighted by the likelihood that an occupation is filled with these qualification levels. The weighting is based on observed (base year) shares of the occupation-qualification matrix. The number of people working with a specific background (qualification) in an occupation determines these weights.

$$p_i = min\left(1, \frac{supply_i}{demand_i}\right)$$
$$IFIOD_j = \frac{\sum_i p_i x_{ij,2010}}{\sum_i x_{ij,2010}}$$
$$0 \leq IFIOD_j \leq 1$$

Where $x_{ij,t-1}$ is the total amount of people in occupation *j* with qualification type *I* in the base year. The indicator gives the relative degree to which difficulties can be expected in fulfilling the demanded occupation. Note that the share p_i is the same for all occupations, as it simply denotes the relative supply to demand of a qualification type. The implicit assumption is that all shortages in qualification types will be felt similarly by all occupations but weighted to the importance of that qualification type for the respective occupation. A value of 1 indicates that no shortages are expected, whereas a (theoretical) 0 would indicate that no demand can be fulfilled. The indicator is usually rank-ordered, in which the quintiles determine the relative level of difficulty to fill a vacancy of that occupation.

Indicator of relative-wage weighted expected imbalance (IRWEI)

The indicator of relative wage weighted expected imbalance (IRWEI) includes one of the key labour market variable into the indicator of future imbalances: the price of labour – wages (Kriechel, 2013; Kriechel, 2014). We use cross-sectional data on wages by occupation and qualification to generate a rank ordering of occupations by qualification. Put simply, occupations in which, on average, the highest wages are paid for a specific qualification level are likely to have adequate supply, even if the overall supply for the qualification level falls short of the total demand. The occupation can be assumed to have first drawing rights on the qualification.

Currently, this is implemented by allowing the upper half of the wage distribution within a qualification level to draw upon half of their demand without a constraint, while the remaining demand for both these and the remaining occupations will have to share the constraint that the remaining demand and supply will generate.

7.3.4. Some thoughts on imbalance indicators

Imbalance indicators should be seen as a tool or a shortcut into interpreting some element of the forecast. They should be used critically, in a sound understanding of how they are derived.

In general, we distinguish between indicators derived from the raw supply-demand comparisons and those derived after we have reconciled supply and demand. The raw supply-demand comparisons, such as all IFIOD based indicators, assume that the separate supply and demand estimates are correct and are not taking any reconciliation into consideration (beyond their own definition). The reconciliation process is likely to be guided by elements such as attractiveness of jobs (thus occupations) which include among other things the wages that a worker can earn (relative to other jobs or occupations). Including wage information can thus make indicators or outcomes of a forecast more realistic. However, in many cases, such information is not available in such a way that would allow the correct forecast of such outcomes. The suggested wage-adjusted indicator is thus only a proxy indicator.

An open element lies in the identification of qualifications. Given the current data situation at the European level, we can only (safely) distinguish between a supply and demand of low, medium, and high qualified. Yet, many occupations require specific fields of qualification, an imbalance that so far has not been included in the project beyond an experimental inclusion of “fields of study” in an experimental setting for which structural data is unfortunately lacking.

Indicators that are using the process of reconciliation or the final results are at first accepting the forecasting outcome – which is after the reconciliation that the occupation model provides – and then try to interpret some aspect of it.

Finally, indicators despite their simple outcome are often hiding many outcomes and assumptions of the entire forecasting process. Thus, results that seem at odds with our current thinking of future developments are usually not caused by an indicator, they are simply shown. Indicators can be used to also find and fine-tune the forecasting process or methodology if integrated into the process.

CHAPTER 8. Summary and conclusions

8.1. Summary

The labour markets of the EU and the individual Member States are undergoing structural changes. The shifts towards a digital and green economy are changing the structure of labour demand, mainly in terms of qualifications required. Ageing means that the labour force must be used more efficiently. Europe cannot afford to waste the potential of the labour force through unemployment and it needs to ensure efficient use of public investment in education. Anticipation of skills supply and demand can provide crucial information about long-term trends and thus support decision-making.

Cedefop’s skills supply and demand forecast is the only European-level exercise based on comparable data. It relies on up-to-date and sophisticated methods and is constantly being developed with the help of key European research institutes.

Anticipation of future skills needs and supply is also taking place in Member States. Cedefop does not intend to compete with or replace efforts at the national level. With this publication, Cedefop intends to present a general overview of the methodology and to inspire those engaged in forecasting exercises or trying to find relevant background material to start their own forecasts.

A consistent and comparable database is a prerequisite for obtaining reliable models and results. EU-LFS data, which provides desegregations by occupation and qualification, and NA data, which provide complete estimates of sectoral employment consistent with sectoral value added and other macroeconomic indicators, are linked.

The Cedefop methodology (Figure 1.1) is based on econometric modelling with certain elements of input-output techniques. The general framework is built on a modular approach, which allows for independent development, fine-tuning and the extension of particular parts in time. The core element is a multisectoral macroeconomic model which produces labour-demand data by sector (E3ME). Its augmented version also produces labour market participation rates for labour supply data. The labour demand by sector is further processed in modules of labour (expansion) demand by occupation and qualification. This forecast also uses a replacement demand module to produce a forecast of those leaving the occupation for different reasons (retirement, mobility, etc.). By combining expansion and replacement demand, the future total job openings can be calculated. Stock and flow modules produce results for labour supply by qualification. A combined stock-flow model would be ideal but is not currently feasible given the lack of data. As labour supply and demand are interrelated, the algorithms to calculate imbalances are not straightforward. Cedefop developed a module on imbalances to deal with this issue.

In Cedefop's forecast, the E3ME provides the links between the labour market and the wider economy. The module is made up of four sets of equations: employment demand, average wages, average hours worked and participation rates. E3ME, which provides a coherent European perspective, allows for the production of different projection scenarios for labour supply and demand.

Projecting future qualification attainments across 32 countries over a period running to 2035 is a demanding task in which various compromises have to be reached in terms of the level of detail and the degree of sophistication at which the modelling can be carried out. Nevertheless, the compromises appear to provide internally consistent results and sensible comparative international findings. There are, however, obviously a number of caveats to be attached to the results in the light of the assumptions and experience gained in the modelling process.

Sample sizes can be very small, both for particular countries and particular qualification levels, especially given the disaggregation by qualification, gender and age. One feature in particular ought to be considered for the future: the low qualification group is both small and projected to be smaller, while the high qualification group is larger and projected to grow larger. As noted in the earlier discussion, this creates small sample size issues in the former which, given the relatively simple methods applied can lead to the need to impose non-negative values of low qualification proportions. On the other hand, the growing size of the high group is no longer so restrictive, which suggests that this group might be usefully disaggregated, particularly as degree and upper degree individuals have taken on increasingly important roles.

Probably the most important issue arising from the modelling as it stands concerns the nature of the data available, which makes so many of the compromises in the analysis necessary. For example, the discussion of the three types of data (cross-sectional, cohort and pseudo-cohort) suggests that none of these is ideal for making projections of likely future developments. While pseudo-cohort appears to give the best route at the present time, the need to adopt a reasonably long period to estimate medium to long-term trends (e.g. 10 years) means that, while some of the educational transitions use relatively recent data (e.g. for 16 year olds), the outcomes for those finishing high level qualifications is dated. It has been argued that neither cross-sectional or cohort data really solve this problem. It is clear that the development and adoption of proper educational accounting systems might potentially help, they require good administrative records, are resource intensive to develop and maintain, and appear to take longer to update than the EU-LFS.

The supply of skills module is based on demographic forecasts and assumptions for individuals to acquire certain level of qualification. Ageing, which is generally spread over European labour markets, will result in decline of younger cohorts in the population and in the labour force. A decline in the number of young people in the labour force is also expected because younger people are staying longer in education than in the past. At the same time, older people usually stay in employment longer due to legislative changes in the age of retirement. Based on such assumptions, the stock of qualifications can be relatively easily derived. It is more complicated to follow the flows among different age categories and age groups important for a more plastic view of the structure of labour supply. A stock–flow model would be ideal for forecasting supply of skills. Producing such a model for the qualification mix of individuals across MSs appears feasible, but is not without problems. The EU-LFS data are now available over a sufficient time period to construct pseudo cohort information that allows modelling of the transition from school to work. But when modelling qualification levels, stock-flow modelling has to deal with the issue of transitions between education pathways. It has been demonstrated that while primarily concentrated among younger individuals, they still cut across a significant age range of at least 16 to 30 years. There are several other minor issues in the modelling process that can be dealt with in a fairly straightforward manner. These include the absence of information on qualifications for those above retirement age and not in employment, and the possible lower bounds for proportions of the population with low qualifications. Other more demanding issues relate to modelling explicitly the effects of emigration and immigration on the qualification mix in different countries. Nevertheless, the present study demonstrates that appropriate data are not available for such modelling at pan- European level.

Changes in the level of employment (expansion demand) disaggregated by occupation and qualification are calculated in two modules (EDMOD and QUALMOD). The projections of employment are derived by mapping projected shares of occupations and qualifications to sectoral employment projections obtained from E3ME. These results incorporate the latest data and, as far as possible, consider latest global developments. In recent years, Cedefop has explored greater level of detail in various data sets, including trials to extend the set of results by more detailed levels of educational attainment and attempts to incorporate fields of study. Given the exploratory nature of the most recent work, not all of the issues and problems have been resolved yet. Although present results are encouraging, so far the approach has not offered a viable alternative to simpler methods of developing projections of changing skill demand patterns based on extrapolative techniques. The conceptual underpinnings of a ‘behavioural’ theory of labour demand, including occupations and education levels, already exist. This theory has its roots in production and cost functions, which have been adapted to estimate employer demand for labour. Subsequent work has been carried out to extend Cedefop’s framework to look at skill-biased technological change (SBTC). The possibility to simultaneously combine estimates of all 81 education/occupation categories was examined. An initial exploration of the data suggests that interesting patterns emerge, but it is still too early to draw conclusions about the success of further work in this area. However, such work would undoubtedly lead to a significant improvement in understanding of recent and current labour demands and employment.

Replacement demand is an important component of the overall labour demand. This component of labour demand represents demand due to people leaving occupations for different reasons who need to be replaced. The cohort component method is used to produce net flows from the labour market. However, the tests of alternative approaches to net flows by occupation estimates can be used only at national level due to the current lack of necessary data sets at European level. We could observe, taking one example for each country, that the flow estimates from panel versus cohort-component approach are quite similar, which leads to conclusion that the cohort component method currently used in Cedefop’s forecast is a good substitute for a methodologically superior panel-based approach if the underlying causes of replacement needs do not need to be identified.

Comparing labour demand and supply to derive potential imbalances is not straightforward. In Cedefop’s skills forecast, supply and demand are calculated separately using comparable data and assumptions. In presenting the implications for imbalances it is essential to emphasise that both the trends in supply (towards a more highly educated workforce) and the trends in demand (towards greater use of such people in employment) are hard to predict precisely. They are also interrelated (supply can to some extent help to generate its own demand, and demand can also generate supply to some degree) but the models are not able to capture these interactions. Several indicators are currently being developed to improve the link between occupational demand and educational supply. Both the indicator based on education shares in occupations and the RAS-based indicator of constraint and measure of change point in the same direction: lower-level occupations will face more adjustment needs and potentially more difficulties in hiring workers. The indicators of imbalances would benefit greatly if a distinction by field of study could be made in education supply.

8.2. Country expert input

Cedefop's conceptual framework, related improvements and results have been developed by a team of experts from leading institutions in economic modelling, labour market analysis and skills anticipation. The uniqueness of Cedefop's approach implies that it is not possible to compare results or methods with similar exercises.

Cedefop has put together a group of individual country experts to ensure the high standard and credibility of the methods used and to validate the plausibility of the results. These experts represent the whole spectrum of professionals starting from practitioners, VET providers through economists and labour market analysts to statisticians and econometricians. Regular workshops, ad hoc discussion groups and informal communication also provide input for further developments and improvements. Without the expert's input and quality checks it would have been difficult to achieve the same outcomes and establish the integrity of the project.

Cedefop is really grateful for each input provided by experts. All other experts and stakeholders are invited to join and provide their own views and suggestions.

References

Bosworth, D., & Wilson, R. (2011). Forecasting skill supply and demand in Europe to 2020: stock flow model and other modelling developments.

Cedefop (2008). Future skill needs in Europe: medium-term forecast; synthesis report.

Cedefop (2009). Future skill supply in Europe: medium-term forecast up to 2020; synthesis report.

Cedefop (2010). Skills supply and demand in Europe: medium-term forecast up to 2020.

Cedefop (2011). What next for skills on the European labour market?

Cedefop (2012). Skills supply and demand in Europe - Methodological framework. Research Paper No 25, Luxembourg: Publications Office of the European Union, 2012

Cedefop (2023) Cedefop Skills Forecast 2023 – Technical Report.

Cörvers , F. (2008). Methodiek arbeidsmarktprognoses en –indicatoren 2007-2012 [Methodology of the labour market forecast and indicators 2007-12].

Engle, E., & Granger, C. (1987). Cointegration and error correction: representation, estimation and testing.

Eurostat. (2013). European system of account (ESA2010).

Hendry, A. (1984). Dynamic specification.

Kriechel, B. (2011). On the flow measure in replacement demand.

Kriechel, B., & Sauermann, J. (2010). Forecasting skill supply and demand in Europe to 2020: replacement demand.

Kriechel, B., & Wilson, R. (2011). Interactions and imbalances.

Kriechel, B., & Wilson, R. (2011). Qualitative evaluation.

Kriechel, B., & Wilson, R. (n.d.). Forecasting skill supply and demand in Europe: imbalances and Mismatches: reconciling the demand for and supply of skills. *2010*.

Livanos, I., & Wilson, R. (2007). Comparing national accounts (NA) and European labour survey (ELFS).

Livanos, I., & Wilson, R. (2007). The Skillsnet project on medium-term forecasts of occupational skill needs in Europe: expansion demand and changes in occupational and qualifications structure.

Livanos, I., & Wilson, R. (2008). Explaining numbers in stock of those acquiring qualifications at a pan-European level.

Livanos, I., & Wilson, R. (2010). Forecasting skill supply and demand for Europe: modelling the demand for skills.

Pollitt, H. (2010). Forecasting skill supply and demand in Europe to 2020: projections of labour supply and demand from the E3ME model.

Shyrock, H., & Siegel, J. (1980). The methods and materials of demography.

Spicer, M., & Reade, J. (2005). Incorporating new EU member economies in multisectoral, multi-region dynamic model: : applying shrinkage estimation to E3ME.

Willems, E., & de Grip, A. (1993). Forecasting replacement demand by occupation by occupation and education.

Wilson, R. (2007). Skills in England 2007: key messages.

Wilson, R., & Bosworth, D. (2010). Medium-term forecast of skill supply in Europe: overview and synthesis.

Annex 1. Acronyms and definitions

Institutions and organisations	
Cedefop	European Centre for the Development of Vocational Training
EU	European Union
Eurostat	Statistical Office of the European Communities
IER	Institute for Employment Research
ILO	International Labour Organisation
OECD	Organisation for Economic Cooperation and Development
ROA	Research Centre for Education and the Labour Market, University of Maastricht
Unesco	United Nations Educational, Scientific and Cultural Organisation

Others	
AMECO	annual macroeconomic database of the European Commission's Directorate-General for Economic and Financial Affairs
BALMOD	module to reconcile skill supply and demand projections
E3ME	energy-environment-economy model of Europe (multisectoral macroeconomic model)
EDMOD	module to produce occupational demand projections (expansion demands)
ESA10	European system of accounts
EU-27	European Union of 27 Member States
EU-27+	European Union of 27 Member States plus Iceland, Norway, Switzerland, Republic of North Macedonia and Turkey
EU-LFS	European Union labour force survey
GDP	gross domestic product
ISCED	International standard classification of education
ISCO	International standard classification of occupations
LFS	labour force survey
LMAR	labour market accounts residuals
NA	national accounts
p.a.	per annum
QMOD	module to produce qualification projections

RDMOD	module to produce projections of replacement demands
StockMOD	module of numbers acquiring qualifications (stocks)

Definitions of terms used

Conceptual framework	The general theoretical and methodological approach to modelling and projecting the demand for and supply of skills.
employment	The number of people in work (headcount) as per the national accounts definition (or the number of jobs in some cases), split by various dimensions, including sector, occupation, gender and highest qualification held.
labour force	The number of people economically active (the sum over the various age ranges of the working age population * the relevant labour market participation rate) which includes employed and unemployed.
population (15+)	Anyone of age 15 or over is classified as part of the population in the context of the model. People over 65 are included in this definition, as these age groups have participation rates greater than zero.
working age population	Anyone of age 15-64 is classified as part of the working age population.
participation or activity rate	The percentage of the population that is either employed or unemployed (ILO definition of labour force). This is differentiated by gender and age group.
qualifications	This term refers to the highest level of education/qualification held by the individual. The ISCED classification is used for this purpose. The most aggregate level distinguishes three main levels of education/qualification: low (ISCED 0-2), medium (ISCED 3-4) and high (ISCED 5-8).
demand	In the context of the model, labour demand is taken to be the same as employment levels (number of jobs available). It does not include (for example) unfilled vacancies.
supply	In the context of the model, labour supply is taken to be the same as the labour force.

Annex 2. Classifications and aggregations used in modelling framework

A1.1. Industries and sectors

Table A. 1 **Aggregation of NACE Rev 2 two and three digit industries to 41 industries**

41-industry [NACE]		NACE rev. 2 [NACE]
1	Agriculture etc [01-03]	Agriculture [01]
		Forestry [02]
		Fishing [03]
2	Mining and quarrying [05-09]	Coal [05]
		Oil and Gas [06]
		Other mining [07-09]
3	Food, Drink & Tobacco [10-12]	Food, Drink & Tobacco [10-12]
4	Textiles, Clothing & Leather [13-15]	Textiles, Clothing & Leather [13-15]
5	Wood, paper, print, publishing [16-18]	Wood and wood products [16]
		Paper and paper products [17]
		Printing [18]
6	Coke & ref petroleum [19]	Manufactured fuels [19]
7	Other chemicals [20]	Other chemicals [20]
8	Pharmaceuticals [21]	Pharmaceuticals [21]
9	Rubber/non-metal min. products [22,23]	Rubber and plastic products [22]
		Non-metallic mineral products [23]
10	Basic metals & metal products [24,25]	Basic metals [24]
		Metal products [25]
11	Optical & electronic equip [26]	Optical & electronic equip [26]
12	Electrical equipment [27]	Electrical equipment [27]
13	Other machinery & equipment [28]	Other machinery & equipment [28]
14	Motor Vehicles [29]	Motor Vehicles [29]
15	Other Transport Equipment [30]	Other Transport Equipment [30]
16	Manufacturing nes [31-33]	Manufacturing nes [31-32]
		Repair & installation of machinery [33]
17	Electricity [35.1]	Electricity [35.1]
18	Gas, steam & air conditioning [35.2,35.3]	Gas, steam & air conditioning [35.2,35.3]
19	Water supply [36-39]	Water supply [36]
		Sewerage and waste [37-39]
20	Construction [41-43]	Construction [41-43]
21	Wholesale and retail trade [45-47]	Trade and repair of motor vehicles [45]
		Other wholesale trade [46]
		Other retail trade [47]

22	Land transport [49]	Land transport [49]
23	Water Transport [50]	Water Transport [50]
24	Air Transport [51]	Air Transport [51]
25	Warehousing and postal services [52,53]	Warehousing [52]
		Postal and courier activities [53]
26	Accommodation & Catering [55,56]	Accommodation & Catering [55,56]
27	Media [58-60]	Publishing activities [58]
		Motion picture and broadcasting activities [59-60]
28	Telecommunications [61]	Telecommunications [61]
29	Computer programming, info serv [62,63]	Computer programming, info serv [62,63]
30	Financial & insurance activ [64-66]	Financial services [64]
		Insurance [65]
		Auxiliary to financial & insurance activities [66]
31	Real estate activities [68]	Real estate activities [68]
32	Legal, account & consulting [69,70]	Legal and accounting [69-70]
33	Architectural & engineering [71]	Architectural & engineering [71]
34	Research & Development [72]	Research & Development [72]
35	Market research & other prof [73-75]	Advertising [73]
		Other professional activities [74-75]
36	Admin. and support services [77-82]	Rental and leasing activities [77]
		Employment activities [78]
		Travel agency, tour operators [79]
		Security and office administrative [80-82]
37	Public administration & defence [84]	Public administration and defence [84]
38	Education [85]	Education [85]
39	Health [86-88]	Human health activities [86]
		Residential care and social work [87-88]
40	Arts and entertainment [90-93]	Arts and entertainment activities [90-92]
		Sports activities [93]
41	Other service activities [94-99]	Membership organisations [94]
		Repair of household goods [95]
		Other personal service activities [96]
		Households as employers of domestic personnel [97]

Table A. 2 **Aggregation of 41 industries to 6 industries**

6 industry [NACE]		41 industry [NACE]
1	Primary sector & utilities [01-03,05-09,35-39]	Agriculture etc [01-03]
		Mining and quarrying [05-09]

		Electricity [35.1]
		Gas, steam & air conditioning [35.2,35.3]
		Water supply [36-39]
2	Manufacturing [10-33]	Food, Drink & Tobacco [10-12]
		Textiles, Clothing & Leather [13-15]
		Wood, paper, print, publishing [16-18]
		Coke & ref petroleum [19]
		Other chemicals [20]
		Pharmaceuticals [21]
		Rubber/non-metal min. products [22,23]
		Basic metals & metal products [24,25]
		Optical & electronic equip [26]
		Electrical equipment [27]
		Other machinery & equipment [28]
		Motor Vehicles [29]
		Other Transport Equipment [30]
		Manufacturing nes [31-33]
3	Construction [41-43]	Construction [41-43]
4	Distribution & transport [45-47,49-53,55,56]	Wholesale and retail trade [45-47]
		Land transport [49]
		Water Transport [50]
		Air Transport [51]
		Warehousing and postal services [52,53]
		Accommodation & Catering [55,56]
5	Business & other services [58-66,68-75,77-82,90-99]	Media [58-60]
		Telecommunications [61]

		Computer programming, info serv [62,63]
		Financial & insurance activ [64-66]
		Real estate activities [68]
		Legal, account & consulting [69,70]
		Architectural & engineering [71]
		Research & Development [72]
		Market research & other prof [73-75]
		Admin. and support services [77-82]
		Arts and entertainment [90-93]
		Other service activities [94-99]
6	Non-marketed services [84-88]	Public administration & defence [84]
		Education [85]
		Health [86-88]

A1.2. Occupations ([15])

Table A. 3 **ISCO-08 occupational classification**

Major group 1: Managers	
11	Chief Executives, Senior Officials and Legislators
12	Administrative and Commercial Managers
13	Production and Specialized Services Managers
14	Hospitality, Retail and Other Services Managers
Major group 2: Professionals	
21	Science and Engineering Professionals
22	Health Professionals
23	Teaching Professionals
24	Business and Administration Professionals
25	Information and Communications Technology Professionals
26	Legal, Social and Cultural Professionals

[15] https://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm

Major group 3: Technicians and associate professionals	
31	Science and Engineering Associate Professionals
32	Health Associate Professionals
33	Business and Administration Associate Professionals
34	Legal, Social, Cultural and Related Associate Professionals
35	Information and Communications Technicians
Major group 4: Clerical support workers	
41	General and Keyboard Clerks
42	Customer Services Clerks
43	Numerical and Material Recording Clerks
44	Other Clerical Support Workers
Major group 5: Services and sales workers	
51	Personal Services Workers
52	Sales Workers
53	Personal Care Workers
54	Protective Services Workers
Major group 6: Skilled Agricultural, Forestry and Fishery Workers	
61	Market-oriented Skilled Agricultural Workers
62	Market-oriented Skilled Forestry, Fishery and Hunting Workers
63	Subsistence Farmers, Fishers, Hunters and Gatherers
Major group 7: Craft and Related Trades Workers	
71	Building and Related Trades Workers (excluding Electricians)
72	Metal, Machinery and Related Trades Workers
73	Handicraft and Printing Workers
74	Electrical and Electronic Trades Workers
75	Food Processing, Woodworking, Garment and Other Craft and Related Trades Workers
Major group 8: Plant and Machine Operators and Assemblers	
81	Stationary Plant and Machine Operators
82	Assemblers
83	Drivers and Mobile Plant Operators
Major group 9: Elementary Occupations	
91	Cleaners and Helpers
92	Agricultural, Forestry and Fishery Labourers
93	Labourers in Mining, Construction, Manufacturing and Transport
94	Food Preparation Assistants
95	Street and Related Sales and Services Workers
96	Refuse Workers and Other Elementary Workers
Major group 0: Armed forces	

A1.3. Qualifications ([16])

Table A. 4 ISCED 2011 (levels of education)classification

Level of qualification	
Low	(Pre)primary and lower secondary (ISCED 0-2)
Medium	Upper and post-secondary (ISCED 3-4)
High	Tertiary (ISCED 5-8)

ISCED 0: Early childhood education

Programmes at ISCED level 0, or early childhood education, are typically designed with a holistic approach to support children's early cognitive, physical, social and emotional development and introduce young children to organized instruction outside of the family context. ISCED level 0 refers to early childhood programmes that have an intentional education component. These programmes aim to develop socio-emotional skills necessary for participation in school and society. They also develop some of the skills needed for academic readiness and prepare children for entry into primary education

ISCED 1: Primary education

Programmes at ISCED level 1, or primary education, are typically designed to provide students with fundamental skills in reading, writing and mathematics (i.e. literacy and numeracy) and establish a solid foundation for learning and understanding core areas of knowledge, personal and social development, in preparation for lower secondary education. It focuses on learning at a basic level of complexity with little, if any, specialisation.

ISCED 2: Lower secondary education

Programmes at ISCED level 2, or lower secondary education, are typically designed to build on the learning outcomes from ISCED level 1. Usually, the aim is to lay the foundation for lifelong learning and human development upon which education systems may then expand further educational opportunities. Some education systems may already offer vocational education programmes at ISCED level 2 to provide individuals with skills relevant to employment.

ISCED 3: Upper secondary education

[16] http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Programmes at ISCED level 3, or upper secondary education, are typically designed to complete secondary education in preparation for tertiary education or provide skills relevant to employment, or both. Programmes at this level offer students more varied, specialised and in-depth instruction than programmes at ISCED level 2. They are more differentiated, with an increased range of options and streams available. Teachers are often highly qualified in the subjects or fields of specialisation they teach, particularly in the higher grades.

ISCED 4: Post-secondary non-tertiary education

Post-secondary non-tertiary education provides learning experiences building on secondary education, preparing for labour market entry as well as tertiary education. It aims at the individual acquisition of knowledge, skills and competencies lower than the level of complexity characteristic of tertiary education. Programmes at ISCED level 4, or post-secondary non-tertiary education, are typically designed to provide individuals who completed ISCED level 3 with nontertiary qualifications required for progression to tertiary education or for employment when their ISCED level 3 qualification does not grant such access. For example, graduates from general ISCED level 3 programmes may choose to complete a non-tertiary vocational qualification; or graduates from vocational ISCED level 3 programmes may choose to increase their level of qualifications or specialise further. The content of ISCED level 4 programmes is not sufficiently complex to be regarded as tertiary education, although it is clearly post-secondary.

ISCED 5: Short-cycle tertiary education

Programmes at ISCED level 5, or short-cycle tertiary education, are often designed to provide participants with professional knowledge, skills and competencies. Typically, they are practically-based, occupationally-specific and prepare students to enter the labour market. However, these programmes may also provide a pathway to other tertiary education programmes. Academic tertiary education programmes below the level of a Bachelor's programme or equivalent are also classified as ISCED level 5.

ISCED 6: Bachelor's or equivalent level

Programmes at ISCED level 6, or Bachelor's or equivalent level, are often designed to provide participants with intermediate academic and/or professional knowledge, skills and competencies, leading to a first degree or equivalent qualification. Programmes at this level are typically theoretically-based but may include practical components and are informed by state of the art research and/or best professional practice. They are traditionally offered by universities and equivalent tertiary educational institutions.

ISCED 7: Master's or equivalent level

Programmes at ISCED level 7, or Master's or equivalent level, are often designed to provide participants with advanced academic and/or professional knowledge, skills and competencies, leading to a second degree or equivalent qualification. Programmes at this level may have a substantial research component but do not yet lead to the award of a doctoral qualification. Typically, programmes at this level are theoretically-based but may include practical components and are informed by state of the art research and/or best professional practice. They are traditionally offered by universities and other tertiary educational institutions.

ISCED 8: Doctoral or equivalent level

Programmes at ISCED level 8, or doctoral or equivalent level, are designed primarily to lead to an advanced research qualification. Programmes at this ISCED level are devoted to advanced study and original research and are typically offered only by research-oriented tertiary educational institutions such as universities. Doctoral programmes exist in both academic and professional fields.

Annex 3. Summary of SORT algorithm functioning

The sorting algorithm at the heart of BALMOD is designed to reconcile the projections from the stock model of supply (numbers available by the three qualification levels) with those from the demand for qualifications model (number of jobs requiring particular qualification levels). The former provides a view of supply-side developments (the overall numbers of people who have acquired qualifications at the three different levels who are actively searching for work), while the latter is more concerned with changing demand for qualifications within occupations (the number of jobs available requiring particular levels of qualifications).

The module also has to deal with differences between the various estimates of employment used in E3ME (based on national accounts and LFS data) and the labour-market accounts residual (LMAR), which arises in part because of such discrepancies but which is also affected by other issues, including measurement error. The main employment measure used in E3ME is a national accounts-based one. This is referred to as unconstrained estimates of employment. All the estimates by sector and occupation are based on this. A second measure, based on LFS information and Eurostat demographic data, is implicit in the modelling of labour supply. This is referred to as supply in employment. The two differ for various reasons, encompassed under the heading of the LMAR. These include:

(a) double jobbing (some have more than one job);
(b) distinction between residence and workplace (many people do not live in the same country as they work; this is especially significant for some small countries such as Luxembourg);
(c) government training and related schemes (which may count as being in the labour force but not as being in employment);
(d) different definitions of unemployment (ILO versus measures of claimants to benefits);
(e) statistical errors (in measures of employment, unemployment and related indicators, including sampling and measurements errors);
(f) other differences due to use of different data sources; treatment of the armed forces and nationals working abroad.

The sorting model uses an iterative RAS procedure to reconcile two sets of estimates of employment, changing the overall qualification shares from the demand for qualifications model (QUALMOD) to match those from the stock model of supply (STOCKMOD). This is done while at the same time maintaining the patterns of occupational deployment and ensuring a plausible pattern of unemployment rates for the different qualification categories. It therefore focuses on the occupations people with different qualifications end up in.

Overall unemployment levels are taken from E3ME. This is taken as exogenous for these purposes. The overall level of unemployment is shared out among qualification categories, based on an extrapolation of patterns from historical LFS data. In the current versions it is assumed that the relative rates of unemployment for the three broad qualification categories are maintained. Checks are made to see that this results in plausible unemployment levels for the three qualification categories. The implied unemployment levels by qualification are then deducted from the overall supply numbers to get the numbers of people in employment by qualification level (supply in employment). The sorting model then reconciles these estimates with the number of jobs available (unconstrained estimates). This is done by altering the shares of people with the three different qualification levels employed within each occupation, until the overall numbers match the numbers of people available.

The final results may provide indications of overqualification or underqualification of people in different occupations, depending on the overall demand-supply balance.

The constraint (matching of numbers by the three qualifications levels) is imposed at the two-digit occupational level. The key dimensions in the SORT routine are:

(a) occupation (41);
(b) qualification level (3);
(c) sector (41).

(note: the results in several summary tables in the imbalances workbooks where this process is undertaken show outcomes for aggregate one digit occupational groups and six broad sectors only).

The sorting model operates for each country separately. There are assumed to be no adjustments through cross-border flows (migration or commuting).

There is then one final step in which the final outcomes from the SORT routine are scaled to match the original E3ME employment totals, to deal with the LMAR discrepancy.