WARWICK INSTITUTE for EMPLOYMENT RESEARCH IER

Using the European LFS to anticipate changing skill needs

R. A. Wilson*

Paper presented at the ISA-RC33 7th International Conference in Naples, September 1-5, 2008
Session Title:
Comparative Analysis of European Microdata from Official Statistics

(Version 1.1)

(* This paper would not have been possible without the contribution of all the other team members, who contributed to the Cedefop funded project upon which it draws, including Frank Cörvers, Ben Gardiner, Ben Kriechel, Hector Pollitt and especially Ilias Livanos who undertook most of the econometric work referred to)

CONTACT DETAILS:

Rob Wilson
Institute for Employment Research
University of Warwick
COVENTRY, CV4 7AL
R.A.Wilson@warwick.ac.uk
Tel: +(44) 2476-523530

Abstract

This paper focuses on the potential for using the European Labour Force Survey (as made available by Eurostat) for developing models and procedures to explain and forecast developments in the demand for and supply of skills. It is based on work carried out on projects for Cedefop. It highlights the problems and difficulties with such an approach and makes some suggestions as to how these might be overcome in future work. It draws directly upon the individual data available from the EU LFS, as well as the published aggregate data that Eurostat make available.

Contents

List of Tables

List of Figures

Summary

The Lisbon agenda has emphasised the need for the regular anticipation of changing skill needs to inform active labour market policies and individuals about the developing situation. These priorities are reflected in the latest official *Employment Guidelines*. Globalisation, technology and demography pose huge challenges, risks and threats to existing jobs, as well as offering many new opportunities. The need for robust labour market information to inform individual's strategic plans and choices is now widely accepted. That these plans and choices should be guided by robust estimates of changing employment patterns, based on regular, systematic and quantitative analysis, is now official policy.

The European LFS provides a potentially invaluable data source, for monitoring such change and developing methods for anticipating future developments. The LFS has become a key EU reference source for estimates of employment (as well as many other key labour market and economic indicators). Non-response rates are typically relatively low and sample sizes are usually relatively large, making the analysis of small subgroups, such as particular skill categories, viable. However, its use also has many difficulties. This paper highlights the potential of such data, as well as a number of problems and pitfalls, including methodological issues, such as data quality, fitness for purpose, cross-national and inter-temporal comparability, and limitations for statistical modelling.

This paper focuses on use of the published and individual microdata provided by national statistical agencies and Eurostat, for all member states. It is based on work undertaken for Cedefop, including the Skillsnet project *Medium-term forecast of occupational skill needs in Europe*. The results of this work highlight the general increase in the demand for skills and, simultaneously, indications of polarization, with significant employment growth in many less skilled jobs (with implications for social exclusion and job quality). It also highlights the dramatic increase in the numbers of people in the EU workforce who have formal qualifications, especially at higher levels.

With regard to the use of the LFS data for such research the paper reaches a number of important conclusions

1. First, the LFS data do enable reasonably robust benchmark projections to be produced at a pan-European level.
2. But, the detailed analysis identifies a number of problems for both demand and supply estimates.
3. Many of these problems can be addressed and dealt with, although this will

require substantial intervention by both Eurostat and the statistical authorities in individual member states. This should be a priority if the ambitions set out in the *Employment Guidelines* are to be met.

4. Some, but not all, of the problems identified can be dealt with by enhancing and enlarging the European LFS. This needs to involve: increased sample size (to provide more robust and precise estimates at a detailed level); additional efforts towards harmonisation (especially with regard to occupational and qualification classifications used, both across countries and over time); and additional questions (to enable analysts to improve understanding of the causal factors at work in the demand for and supply of skills).
5. However, not all of the problems can be dealt with by improving the LFS. There are also a number of gaps in the existing European statistical infrastructure that can only be filled by ***New*** surveys, including a survey of employers to measure much more precisely changing patterns in the demand for skills. A pan-European survey focused on **employers** rather than households should be a priority.
6. There are lessons to be learned in this regard from best practice elsewhere in the world. Such a development would represent a major investment, with substantial cost. However, the cost of ignorance about such matters could be even greater, if Europe fails to take such matters as seriously as its competitors across the rest of the world.

Keywords: EU-Labour Force Survey (EU-LFS), comparative research, skill, occupation, occupational projections, anticipation of changing skill needs, quantitative modelling

1. Introduction

This paper discusses the potential value of European Labour Force Survey data for monitoring changes in the structure of employment over time, with a view to informing policy makers and other labour market participants about how the future demand for and supply of skills is likely to develop.

The Lisbon agenda has emphasised the need for the regular anticipation of changing skill needs to inform active labour market policies and individuals about the developing situation. These priorities have now been recognised in the official EU *Employment Guidelines.*[1] Additional resources are being deployed at pan-European level to monitor historical changes and develop projections of how things may change in the future.[2]

Globalisation, technology and demography pose huge challenges to the EU. The increase in competitive pressures as countries such as China and India have increased their involvement in global markets, and the increasing ease with which newly developing countries can compete on a level playing field with the more established countries, have increased the risks and threats to many existing jobs. However they also open up new opportunities as these emergent economies demand new goods and services as their populations get richer.

A number of authors have argued that such developments strengthen the case for a regular, systematic and consistent assessment of changing skill demands and supplies across Europe, (see for example Wilson (2007)). Of course the future cannot be predicted precisely, but the need for robust information to guide strategic plans and choices is now widely accepted. These plans and choices can influence the future path taken by the economy and labour market. It is also now accepted that they should be guided by robust estimates of changing employment patterns, based on regular, systematic and quantitative analysis. This has now been formally recognised in the *Employment Guidelines*.

In order to help meet these objectives, Cedefop have funded projects to develop a "*Medium-term forecast of occupational skill needs in Europe*" as well as "*Medium-term forecasts of supply skills in Europe*". These two projects have been conducted by the Warwick Institute for Employment Research (IER), in collaboration with Cambridge Econometrics (CE) and the Research Centre for Education and the Labour Market (ROA) in Maastricht.

[1] See http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/lsa/101012.pdf

[2] See for example the recent call for tender from Cedefop for a 4 year Framework agreement to produce regular quantiative projections.

The prime objective of this paper is to assess the way in which European Union Labour Force Survey (EU LFS) data have been used to help meet these objectives. In particular it focuses on:

- the use of the LFS to estimate changing occupational employment shares within sectors; and
- trends in the stocks of people categorized by the highest qualification held.

The European LFS provides a potentially invaluable source for monitoring change in such employment patterns. It has become a key EU reference source for estimates of employment (as well as many other key labour market and economic indicators). In many countries it is now being used as a key input into their National Accounts. Response rates in the LFS are usually good and sample sizes are relatively large. This makes the data suitable for estimating the relative size (in terms of employment) of small subgroups such as particular skill categories (measured by occupation or qualifications) within sectors.

The paper briefly summarises some of the results from the Cedefop funded research, highlighting both the value of the LFS for identifying key trends, but also its limitations. Its use for such purposes also has many difficulties. This paper considers the strengths and weaknesses of the existing LFS data, as made available by Eurostat. It highlights a number of problems and pitfalls in using it for measuring changes in employment structure across Europe in a consistent fashion. These include methodological issues such as:

- data quality;
- fitness for purpose;
- cross-national and inter-temporal comparability; and
- limitations for statistical/econometric modelling.

The paper then makes various suggestions as to how matters might be improved.

The structure of the remainder of this paper is as follows:

Section 2 briefly summarises the key research questions the work is addressing;

Section 3 reviews the LFS generally, and discusses more specifically its strengths and weakness for these purposes;

Section 4 presents some illustrative results (highlighting a number of problem areas);

Section 5 considers the case for a new pan-European survey focused upon employers rather than households, to complement the LFS and provide more robust data on the demand for skills; finally

Section 6 concludes by discussing the main implications, including improving the existing LFS data and the need for a new survey.

2. Key Research Questions: Methods and approaches to modelling skill demand and supply

2.1 Two key Questions

There are two key questions that the research for Cedefop aims to address:

1. What are the likely changes in the patterns of demand for skills as measured by occupational employment and the qualifications of those employed?
2. What are the likely changes in the supply of skills, as measured by the formal qualifications people hold?

The next two sub-sections briefly expand on what these questions entail, focussing in particular on the kinds of data required to answer them.

Livanos and Wilson (2007b and 2008) review the recent literature on the demand for and supply of skills, with special emphasis on the quantitative modelling of changes in occupational and qualification employment structures. They summarise the econometric specifications and data used in previous research. These insights were taken into consideration in developing the framework for producing both demand and supply projections that constitute a "best practice" approach at a pan-European level (Wilson *et al.* (2008). In addition to presenting various simple extrapolative approaches, they explore methods for estimating occupational shares using a multinomial logistic regression. This same general approach is also used to analyse qualification shares within occupations, as well as the supply of people with formal qualifications.

2.2 The Demand for skills

The review of best practice, world-wide in Wilson and Livanos (2007) suggests that most occupational projections are based on two key elements:

- a multi-sectoral macroeconomic model that generates prospective views of employment levels within each sector; and
- a detailed analysis of occupational employment shares within sectors;

Projections of occupational employment are obtained by applying the projected occupational shares within sectors to projected estimates of sectoral employment from the macroeconomic model.

The review of the minimum data requirements for best practice occupational forecasts (often loosely referred to as demand for skills) suggest that the key data requirement is therefore detailed and robust information on employment by occupation (and qualification)

within sectors. By analysing trends in such employment patterns over time, and projecting these into the future, such information can (in combination with the output from the macroeconomic model) be used to generate employment projections. Such methods are already used in many EU countries, and also in many countries outside the EU (including the USA, Canada and Australia).

In the USA, for example, the occupational projections of the Bureau of Labour Statistics (BLS) have been published for many years in the Monthly Labor Review (e.g. Hecher, 2005). Based on recent correspondence, their overall approach has not changed much since it was set out in BLS (1997). This last report sets out the series of steps used by the BLS in order to determine the shifts in occupational shares expected in the future. The key is a detailed review of the historical data based on the Occupational Employment Statistics survey (OES) and sectoral employment projections based on a multisectoral macroeconomic model (based around a Leontieff input-output analysis). Projections of occupational employment shares are produced by a combination of fairly simple extrapolations of historical trends in occupational patterns within sectors, combined with analytical studies of specific industries and occupations, technological change, a variety of other economic data, and judgements about how the patterns will change in the future.

In most countries the analysis of the occupational shares is quite simplistic, involving simple trend extrapolation rather than sophisticated behavioural analysis. In contrast, the multi-sectoral models used to generate the sectoral projections are usually very complex, drawing upon leading edge econometric and economic theory and practice. The difference between the two elements is largely due to differences in the quality of data available to the analyst.

Macroeconomic models have benefited from many decades of investment in statistical infrastructure related to the national accounts and input-output analysis. Because of these prior investments in data, the macroeconomist has long time series of data on a consistent basis to work with. In contrast, labour market analysts, focussing on occupations and qualifications, are much less well served, especially at a pan-European level. In most cases there is only a very short time series of data available. This severely limits the kind of time series models that can be developed.

Gradually, however, the data situation is improving. As a result, various efforts have been made worldwide, to develop more sophisticated approaches to modelling changing occupational employment patterns, using econometric methods that capture the links between employment and factor prices suggested by economic theory. For individual countries such as the UK the development of longer time series of data on occupational

employment have encouraged more sophisticated analysis of changing occupational patterns within sectors and at an aggregate level.[3] However, in general, these efforts have met with mixed success. In most cases a much simpler approach, based on "time series" methods, is adopted. The shares of those employed in different occupations or skills categories is simply related to past trends, coupled with a large dose of judgement.

The situation has also been improving on a pan-European level, as Eurostat have striven to develop a comparable set of LFS data covering all member states. This has resulted in the initiatives by Cedefop to explore the feasibility of using such data to produce regular, consistent and systematic assessments of changing skill needs across Europe. In principle, using the EU LFS data, occupational share coefficients for each sector can now be calculated for quite a long period. Various methods can then be used to model these historical patterns (including linear or non-linear regression analysis) and then projections developed. Such an exercise has now been completed at a European level in the Cedefop project (Wilson *et al.* 2008). The present paper reflects on what lessons can be learned from this process from a methodological perspective, and what needs to be done to improve the data and methodology in order to obtain more robust estimates in the future.

Efforts to estimate a consistent behavioural model of changing occupational employment structure at a pan-European level, have so far failed to produce robust results, primarily due to data limitations and constraints (especially the very limited number of time series observations available on a consistent basis). The most robust model developed for the Cedefop project uses time as the only regressor. Time can be regarded as a proxy for technological change and other highly trended variables that may influence occupational or qualification structures.

Economic theory, and the few "best practice" studies conducted in other countries referred to above, suggest that, given longer and more consistent time series and better quality data in general, more sophisticated behavioural explanations are possible. These can take into account substitution effects, as well as recognising the impact of other potential drivers of changing occupational structures.

[3] Briscoe and Wilson (2003) used annual data from successive UK Labour Force Surveys to model occupational trends over the period 1981 to 1999. They developed time series models for some nine occupational groups across 17 industrial sectors. The variables they used included; output, wages, unemployment, export and import shares, which were combined with different technology (trend) and interactive dummy terms to identify the most significant determinants of the changing demand for occupational skills. Similar models have been developed in a number of other countries, some of which attempt to develop behavioural models that take into account various economic variables such as international trade and technological change, but these have been the exception rather than the rule. Moreover, in most (if not all) countries which have conducted similar work, experts' expectations and judgments still play an important part in predicting the future occupational shares.

2.3 The supply of skills

The second key question addressed in the Cedefop work, is What is the likely change in the number of people in the labour force holding different formal qualifications?

The review in Livanos and Wilson (2008) emphasises that educational attainment is affected by a number of factors. The most well established approach is human capital theory (Becker 1964, 1993). Education is seen as an investment, which is expected to have positive financial returns. Individuals invest in education until the expected returns for an additional year's investment equals the expected costs. There are of course a number of other reasons why individuals acquire education (e.g. social status, family considerations, personal interest in a subject etc.). There are also competing theoretical explanations for why it is beneficial to acquire formal qualifications or credentials such as the screening hypothesis (Layard and Psacharopoulos 1974) and credentialism (Collins 1979). These theories highlight the value of qualifications as a signal of an individual's innate ability rather than as a measure of improved productivity as a consequence of the education received.

The key data requirements here are time series estimates of numbers of people with qualifications, as well as information on the flows of those acquiring qualifications at different levels. Again this can be done using the European LFS (although the Cedefop project also draws upon other data sets for flows including the UNESCO-OECD-Eurostat (UOE) database of education statistics, compiled from national administrative sources as reported by Ministries of Education or National Statistical Offices).

LFS data can in principle provide information on the numbers in the total population and the labour force qualified at quite a detailed level, based on the ISCED system of classification (see Annex A). The ISCED based variables in the LFS provide up to 19 different levels, as well as some information on subject of qualification for higher level qualifications.

In principle, the LFS also provides information on many of the indictors that might be regarded as influencing supply decisions at the individual level. It therefore appears to be ideally suited for this purpose. In practice, things are not so simple as the discussion in the following section makes clear.

3. Strengths and weaknesses of the European LFS as a source of data on the demand for and supply of skills

3.1 Estimates of occupational employment structure

The EU LFS, conducted in a common fashion across the EU, provides a potentially invaluable source of information on employment by industry and occupation (as well as qualification). In many member States, LFS data have become the key source for employment estimates. They have the advantage of being conducted on a much more frequent basis than a typical Census, which (up until the development of the LFS) used to be the only way of comparing occupational patterns across member states.

From the perspective of developing a pan-European data set, the Eurostat LFS dataset adopts a much more standardised set of questions and systems of classification across countries than is possible by comparing Censuses. While there are still a number of important differences across countries, the LFS provides (in principle at least) a broadly consistent set of data which can be used for comparing occupational and qualification employment patterns across all member states. Such a data set can be used to begin to develop the quantitative projections which the Lisbon Agenda calls for.

The project funded by Cedefop (Wilson *et al.* 2008) utilised aggregate data from the EU-LFS, as published by Eurostat, as well as the micro dataset. These data cover 27 countries for the period 1983-2006, and contain information on employment by country, gender, industry (41 industries), occupation (27 occupations), qualification (3 level qualifications). These data enable the measurement of the changing pattern of skill demand, in theory, harmonised to a common classification. The project also utilised the LFS micro-dataset on individuals.[4] For a full description of the list of variables included in the LFS see European Communities (2003).

Occupational employment patterns are only one way of measuring skill. The types of formal qualifications typically required are also important. Many countries include a qualification dimension in their employment projections. It is possible to create employment matrices by occupation cross-classified by qualification from LFS data.

[4] The LFS micro-data contains sectoral information but only at the 1 digit level (16 sectors), due to the anonymization procedures adopted by Eurostat. This is a major drawback as it restricts the analysis below the desired level of disaggregation. The micro dataset provided by Eurostat for the Cedefop project also did not include information on pay, although this is now becoming available for at least some countries. For these reasons, the micro dataset was not ideally suited for the analysis, which was focused on trying to examine developments at the most detailed level possible by sector and occupation.

3.2 Replacement Demands

LFS data are also useful for measuring so called replacement demands. These focus on the need to replace those individuals flowing out of the workforce for retirement or other reasons. Estimating replacement demand is not straightforward and is quite sensitive to both the methods and the data sources used (for a review see Fox and Comerford, 2006). Most previous work has tended to focus on what might be called "permanent or semi-permanent" withdrawals from the employed workforce. These include:

- Mortality;
- Retirement (and other reasons for leaving the workforce, including family formation);
- Emigration;
- Inter-occupational mobility.

To estimate such replacement demands, information on the age and gender structure of employment is usually required because many of the flows, especially retirements and mortality, are age and gender specific. Age structures also vary significantly by occupation. Differences in age structure across occupations can clearly influence exits, with more older people retiring, but more younger people changing occupations. Age structure also affects mortality.

Using the LFS, it is possible to analyse the demographic composition of each occupation. This makes it possible, in principle, to estimate specific rates of retirement and mortality for each occupational class. LFS data can also be used for making estimates of rates of outflow. However, there are problems in obtaining robust data, since even for fairly broad age categories the samples are quite small in many cases, and the sampling errors large.

3.3 LFS data for modelling the supply of qualifications

Ideally the analysis of labour supply requires micro-data on individuals. In principle, the European LFS looks well suited for this. It is a sample survey of households and is conducted by the national statistical agencies on a quarterly basis. The data available cover the period 1983-2006 for 27 countries. However, in practice, the data series are often incomplete, and the period for which consistent data are available is much shorter and varies across countries.

The LFS data provide information at the level of the individual regarding:

- general characteristics (age, gender, nationality etc);
- the individual's labour market activity over the reference week (whether they are employed/unemployed etc);
- the individual's main job (occupation, sector, employment status, hours of work etc);
- flexible working patterns (evening work, overnight etc);
- similar information on second and other additional jobs held;
- information on the individual's previous employment (if unemployed);
- methods of looking for a job; and,
- educational attainment.

Although the LFS is a household survey, the Eurostat version available for analysis in the Cedefop project excluded a number of key variables related to the household, such as family income, social status, parent's education and occupation, as well as the income of the individual.[5] This is quite restrictive since, as indicated by the review above, such variables are often utilized for investigating educational attainment. In terms of ideal data requirements, the present data available are therefore lacking.

3.4 Problems in using LFS data to measure occupational employment structure

As described in more detail in Livanos and Wilson (2007a), the LFS is not without its faults when trying to measure changes in occupational employment structure. The LFS data from Eurostat are sample data, based on a survey of households. A key problem is that the kind of analysis described in Section 3.2 requires very detailed data by sector and occupation. Slicing the LFS data across multiple dimensions simultaneously results in problems of small or empty cell sizes.

These problems are even more serious when it comes to making estimates of replacement demand, which asks even more from the data. Nevertheless LFS data are in many cases the only data available. In practice, the sparse nature of the industry by occupation employment data, in both the published and the micro dataset, mean that the LFS data set was not as quite as valuable as was originally the hope when the Cedefop project was initiated.

There are a number of particular issues that need to be highlighted regarding both the microdata set and the published Eurostat data. For many countries there is only a quite short consistent data series. In many instances there are breaks and gaps in the data. In addition, as can be seen in the graphs presented below in Section 4, there is often

[5] Subsequently the latest version of the data provided by Eurostat now has a variable called ***incmon*** that refers to net pay from main job. This variable provides the decile that the individual is in. However, this variable is optional and is currently available for only a limited number of countries.

considerable volatility in the data. This reflects statistical noise (due in part to insufficient sample size). The number of respondents within a particular cell is often very low. This means that the estimates of occupational structure within sectors are not always precise or robust. In addition there are often changes in survey methods and systems of classification, which result in discontinuities over time rather than real change.

A further problem with using LFS data is that there are often some significant differences compared with estimates of employment based on National Accounts. There are many differences between the two sets of estimates. Annex C provides a more detailed discussion of the differences between the two sets of estimates. In many respects estimates of employment based on National Accounts are more robust and more compatible with other economic indicators. For example in the LFS somebody who works just 1 hour in the reference period is counted as employed whereas national accounts based measures usually imply a more substantial commitment (often using full time equivalent concepts). However, LFS estimates have now become well established in analytical and political discourse (e.g. they are used to measure European benchmarks and indicators), with the result that some people are more familiar with these data than those based on the National Accounts.

Another key problem for making precise estimates of occupational structure is that the LFS relies on the respondent's own assessment of occupation (or if the individual concerned is absent at the time of the interview a proxy response is used from a member of the household actually present). Self assessment may be biased, while proxy responses raise concerns about accuracy of the assignment to particular occupational categories. Other problems arise due to double jobbing (many people have more than one job) and the fact that, as noted above, the LFS definition of employment includes anyone working for just a single hour during the reference period.

There are a number of other issues relating to data comparability and quality which emerged in the course of the Cedefop project. These are discussed in more detail in Livanos and Wilson (2007a and 2007c). They include variations in estimates in different vintages of both LFS and NA data sets, as well as differences between the published LFS data and those available in the micro datasets provided by Eurostat for the econometric analysis. Many of these data series are, in principle, from the same source but differ as a result of being collected, collated and published at different times. As a consequence, there are a number of different versions of each of the data sets available at any one time and considerable care is required when making comparisons to avoid spurious comparisons. Annex B highlights some of these problems.

Finally there are still some major concerns about how well harmonised the data are across countries. Casual inspection suggests some very significant differences in the importance of some occupational group across countries, as well as some significant differences in qualifications. This may reflect some remaining differences with the way jobs are classified across countries, as well as difficulties in fitting some national qualifications into the ISCO and ISCED classifications, despite attempts to harmonise the systems of classification. For example, there appear to be some general problems with managers. There appear to be many more managers in the UK (and in one or two other countries) than in most other countries. Elias and Birch (2008) state:

a comparison of occupational structures in 2003 presented in Employment in Europe 2004 shows that, for the proportion of employment in ISCO 1988 Major Group 1, nearly all EU countries and the US show a deviation from the EU average of +/- 2.5 percentage points . However, the UK and Ireland stand out by showing a deviation of +6 percentage points from the EU average. Given the strong similarities between the UK economy, many of our EU partners and the US, it is unlikely that this reflects major differences in the organisation and structure of work in the UK and most probably relates to the classification methods and procedures embodied within occupational data classified to SOC 2000 Major Group 1.

This kind of problem has been recognised at both national and international level and steps are being taken to try to harmonise things better.[6]

3.5 Problems in measuring qualifications

There are some more specific problems for the data on qualifications. For instance, for a small number of countries (e.g. UK, DE) there are no data on levels of qualifications for 1998. More importantly there are question marks against the comparability of some of the data across countries (the treatment of medium level qualifications in the UK for example).[7] These problems with the data create a very imprecise picture about the historical levels of employment, and this limits what can be achieved in modelling, and therefore affects any projections. These issues are discussed in greater detail in the following section.

[6] See for example Elias and Birch (2008) in the case of the UK.

[7] See Section 4 for further details.

4. Illustrative Results: Implications of the data problems

4.1 Historical trends in occupational employment

This section focuses on some of the problems caused by the limitations of the LFS data, focussing upon occupational shares. Before highlighting these, it is important to emphasise that the results on projections of occupational and qualifications shares reported in Wilson and Livanos (2007b), suggest that the underlying trends in occupational employment structure are generally quite robust and not too sensitive to alternative specifications. In most cases, the specifications deployed suggest a strong increase in employment for high-skilled occupations (e.g. managers, professionals and associate professionals) and simultaneously a decline for many less skilled ones (e.g. office clerks, and machine operatives). There are some exceptions to this broad rule, with some indications of a polarisation of demand for skills – in particular there appears to be a growing number of jobs (especially in the service sector) that require only elementary skills.

Of course, there are many detailed variations across countries and industries. While all methods paint a broadly similar picture regarding the future of occupational and qualification employment, nevertheless, the exact results and shares using the different methods and specifications show considerable variation, especially for the detailed industries and occupations. At this more detailed level the results also begin to betray some of the limitations of the LFS data. This section presents some examples of the kinds of problems this poses.

The following figures are for selected countries focussing mainly on **ALL** industries and services. The results for some countries and for individual sectors, especially the smaller ones show much greater volatility and even more gaps and discontinuities, some of which have implications for the projections. The same is true (if anything even more so) of the data disaggregated by qualification. Table 4.1 provides a summary of the situation country by country, identifying the most obvious breaks.

Figure 4.1 is for the UK and represents what appears to be an “ideal” case, with no major problems. However closer inspection reveals a number of difficulties, including quite a marked discontinuity in 2000/2001 linked to a change in official classification of occupations. While in the UK efforts have been made by researchers to adjust the earlier data based on SOC 1990 classification to match the later SOC 2000 classification, such a harmonised time series data set has never been officially produced. The data passed on to Eurostat therefore include a major discontinuity.

Similar problems appear to exist for many other countries although these are not always flagged up to potential users. The article by Eurostat (2007) highlights many general discontinuities in the LFS but does not mention breaks due to changes in occupational or qualification classification. Yet this appears to be a serious problem in many cases in addition to the UK one noted above.

For example, the data for France exhibits another major discontinuity in 1994/95. For this reason the data for this country, and others in a similar position, can only be used for the years after the obvious statistical break. Ireland exhibits a similar break but rather later (In 1998/99). The data set can again be truncated, but this leaves a much shorter period to establish any meaningful trends. In the case of Italy there is a much more recent discontinuity. This leaves too few observations to discard the years in the data set from before the break. This illustrates a problem common to a number of other countries, including Austria and Poland.

For many countries data are only available for recent years. In some cases there are also missing years at the end (e.g. in the data set used for the Cedefop project Luxembourg and Switzerland were missing data for 2006). Such short time series cannot provide very reliable guides as to current trends, although as time goes by, and data are added for more years, this type of problem will gradually disappear.

Figure 4.2 illustrates that although at an aggregate level, across all industries, the UK may appear to have some reasonably stable trends, at a more detailed level this may be far from the truth. The diagram shows the results for “Coal mining”. This sector now employs just a few people in the UK. As a consequence the LFS estimates are erratic and provide little guidance on likely future trends. The one consoling thought here is that such sectors are by definition not very important (at least in terms of total numbers employed).

Table 4.1: Major Breaks and discontinuities in the LFS data on occupations

Denmark	no obvious break	
Germany	discontinuity in 2001	discontinuity due to re-unification
Estonia	truncated 1997	
Greece	no obvious break	
Spain	no obvious break	
France	discontinuity 1994/95	
Ireland	discontinuity 1997-1999	
Italy	discontinuity 2003/04	
Cyprus	truncated 1999	
Latvia	truncated 1998	
Lithuania	truncated 1998	
Luxembourg	truncated after 2006	
Hungary	truncated 1997	
Malta	truncated 2000	
Netherlands	truncated 1996	
Norway	truncated 1996	
Austria	truncated 1995 discontinuity 2003/04	
Poland	truncated 1997 discontinuity 2004/05	
Portugal	discontinuity 1997/98	
Switzerland	truncated 1996 & 2006 discontinuities up to 2001	
Slovenia	truncated 1996	
Slovakia	truncated 1998	
Finland	truncated 1997 discontinuity 2001/02	
Sweden	truncated 1997	
United Kingdom	discontinuity 2000/01	

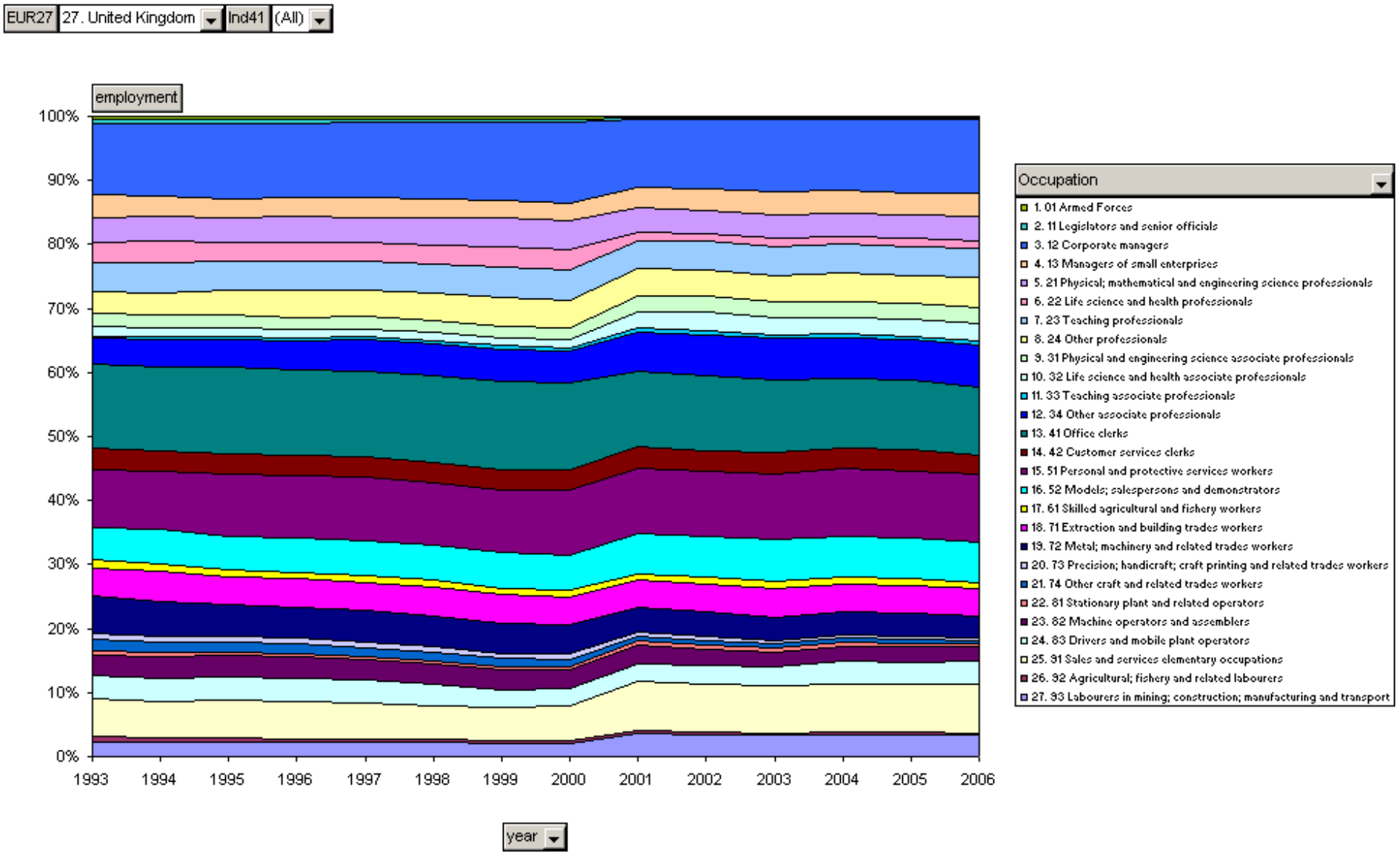


Figure 4.1: Changing Occupational shares- United Kingdom

Source: IER estimates, based on Eurostata LFS data (N:\Projects\Eforecast\Workbooks\pivotTables LogExtraOccPT.xls)

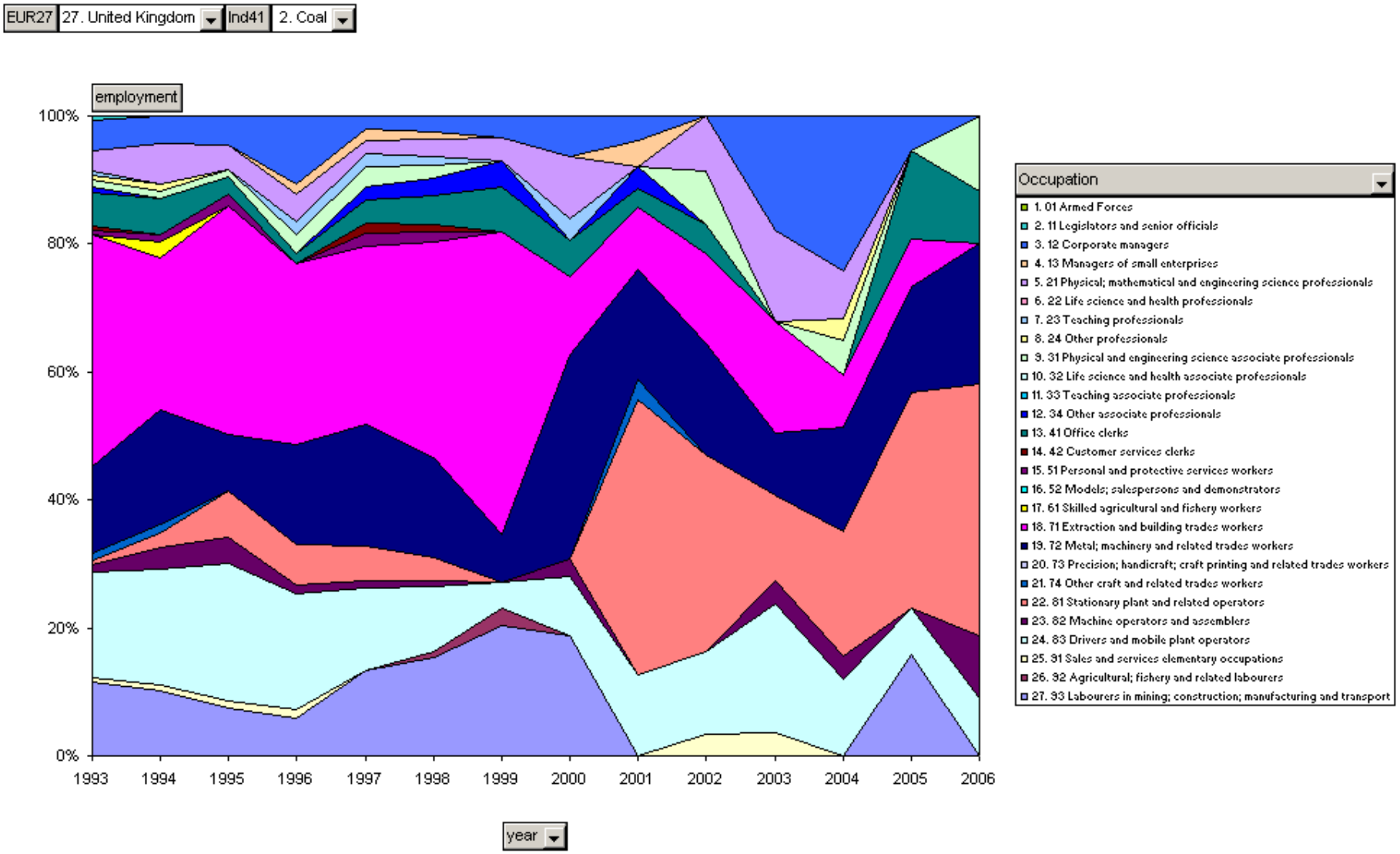


Figure 4.2: Changing Occupational shares- UK, Coal Industry

Source: IER estimates, based on Eurostata LFS data

4.2 Implications for modelling and projections of occupational structure

Based on the review of the literature in Wilson and Livanos (2007b) an ideal specification for modelling (and the projecting) occupational employment structure is shown in equation (1):

S_{ijt} = f(Year, Technology$_{ijt,}$ Trade$_{ijt}$, Wage$_{ijt}$, Output$_{ijt}$, Unemployment$_{jt}$, X_{ijt}) (1)

where S_{ijt} denotes the share of employment by occupation (j) within industry (i) at time t, and X denotes a vector of other employment characteristics such as country of residence and gender. This represents the long term relationship between the demand for skills and various key drivers. In practice it can be anticipated that there will be various lags and adjustments which require the exploration of the dynamics of this relationship.

Given the short time series available in the LFS, the ability to include a full set of economic indicators as explanatory variables is severely limited, resulting in the use of a number of simpler specifications of the general form shown in equation (2)

S_{ijt} = F(time) (2)

These range from simple extrapolation between fixed points, to various methods based on 'line fitting'. The latter includes fitting:

- a linear trend, [S = a + b*Time] (3)
- a log linear trend [Ln(S) = a + b*Time] (4) or
- a logistic equation [Ln(S/(1-S)) = a + b*Time] (5)

Two main methods of analysis were adopted. The first is based on analysis of the aggregate employment shares in the published LFS data, and adopting specifications as in equation (2). These include a number of variations including linear, semi log and logistic forms. Such equations can be regarded as simplified versions of equation (1), where time is used as a proxy for technological change.

The second method of analysing changes in employment structure involved estimating a probability model, using individual data. The propensity to be employed in a particular occupation is modelled using a multinomial logistic regression framework, based on pooled cross sectional data from the Eurostat/LFS microdata. (For details of the mlogit approach see Annex D). An important advantage of the mlogit method is that the sum of probabilities is constrained by the model to add up to one (i.e. shares of employment across occupations/qualifications sum to 100 percent).

The most robust results were obtained using simple country datasets that included just information on time, occupation, qualification, and sector. Mlogit analyses were also conducted in order to estimate the shares of qualifications within each occupation, within every industry in each country.

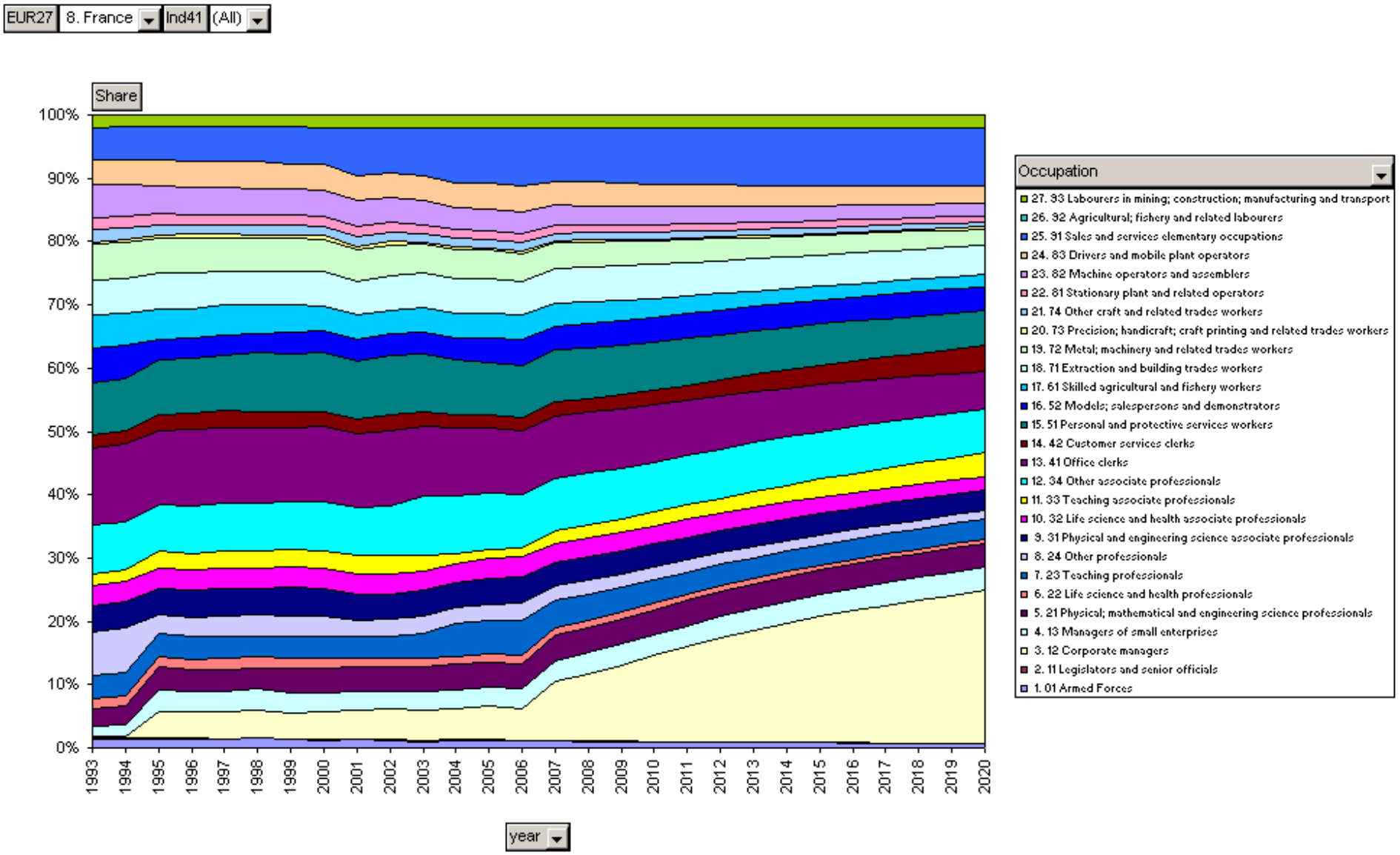


Figure 4.3: Changing Occupational shares- France, projections for all industries

Source: IER estimates, based on Eurostata LFS data, **(**N:\Projects\Eforecast\Workbooks\pivotTables\old LogExtraOccPT.1.xls)

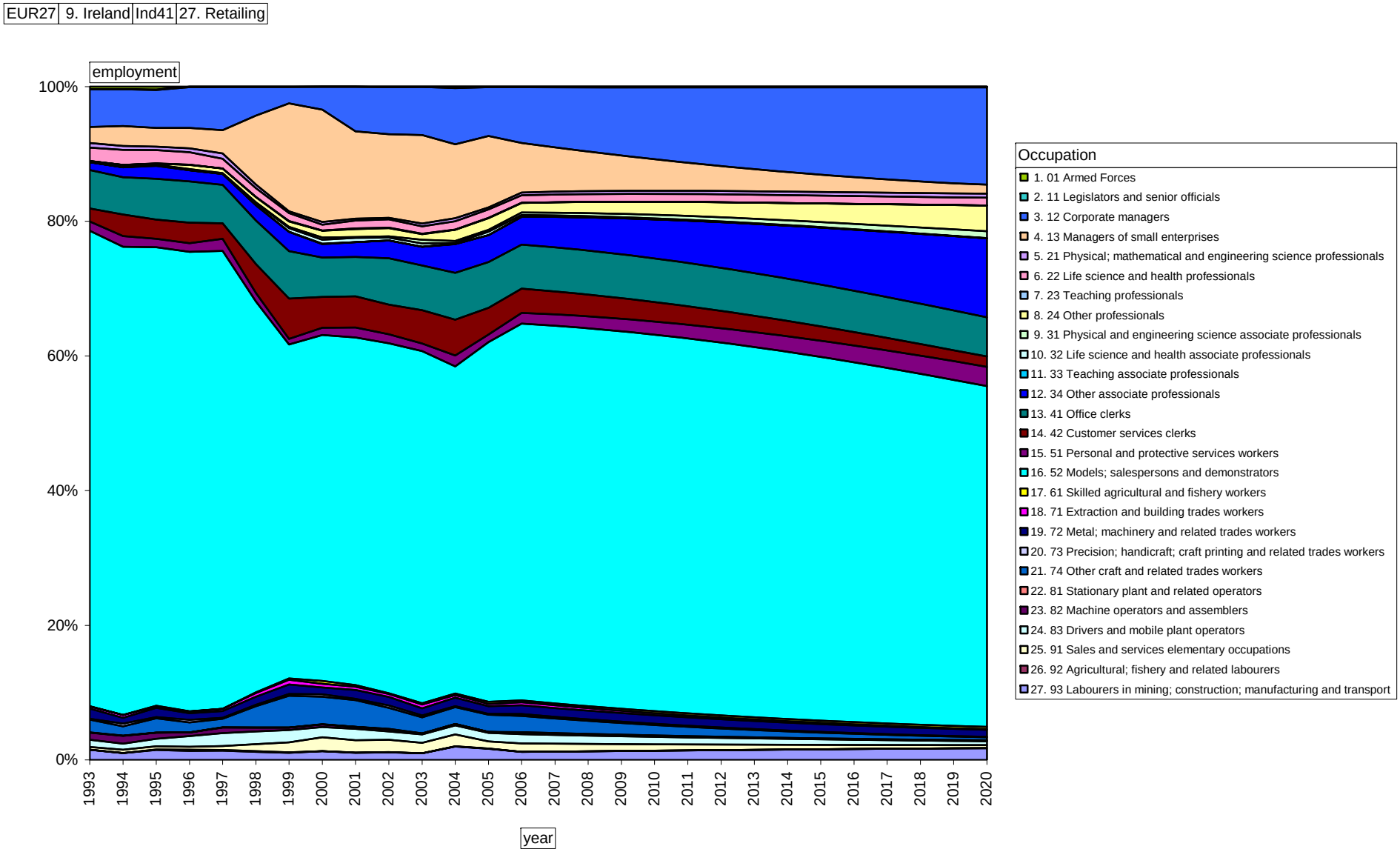


Figure 4.4 Changing Occupational shares- Ireland, projections for retailing

Source: IER estimates, based on Eurostata LFS data, **(**N:\Projects\Eforecast\Workbooks\pivotTables\old LogExtraOccPT.1.xls)

Attempts to estimate more general models that allowed for greater behavioural content by including additional economic variables in the specification failed to establish any consistent and robust results. This does not mean that such models may not be relevant but that it proved impossible to find suitable specifications given the data currently available. The final models were estimated for each of the 41 sectors separately and each of the Eur27[8] countries.

The results of the multinomial logistic regression analysis were then used in order to make predictions of future occupational shares in each sector and in each country (2-digit level), as well as shares of qualifications for each occupation, within the 41 industries used in the macroeconomic model.

The mlogit models estimated were quite robust. In general, over 90% of all coefficients in the analysis were statistically significant at the 95% level. The robustness of the estimated models provides a reasonably sound foundation for making the predictions for the shares of occupations and qualifications. However, the short length of the time series involved and other difficulties with the data cause problems in some cases.

Figure 4.3, shows some of the possible implications of the type of data discontinuities highlighted when attempts are made to develop projections using these models (in this case for France). Using the full time series, the results indicate an unlikely explosion in the number of **1.2: Corporate managers**. This is a reflection of a change in classification rather than reality. It disappears if the period used for modelling the trend starts in 1995 rather than 1993. Ideally the data series for such countries should be replaced by a more consistent one covering all years, but this requires intensive work to identify the cause of the discontinuity and to correct it.

Figure 4.4 illustrates another common problem, in this case for Ireland, in retailing. Sample sizes in the LFS are much larger here but this does not avoid what appear to be very erratic changes over time. This can result in strange projections using extrapolative methods of whatever kind. Note the very rapid growth projected for **3.4: Other associate professionals** and the projected collapse of employment in **1.3: Managers of small enterprises**.

Unfortunately this kind of problem is not exceptional. Some are related to discontinuities in the basic Eurostat LFS data. Others are more a function of the lack of precision and robustness of these data (with many cells being empty or containing estimates with large margins of error). Despite considerable effort it proved impossible within the Cedefop project to develop models or rules to avoid entirely these kinds of problems. There is no single model or method that avoids all such difficulties. A purely technocratic solution is not possible and considerable judgement is needed to chose the “best” method and projection. This suggests that there is a need for greater

[8] Eur27 refers to EU25 (as of December 2006) plus Norway and Switzerland.

and more detailed scrutiny of the basic data, especially by national statistical authorities BEFORE supplying the data to Eurostat.

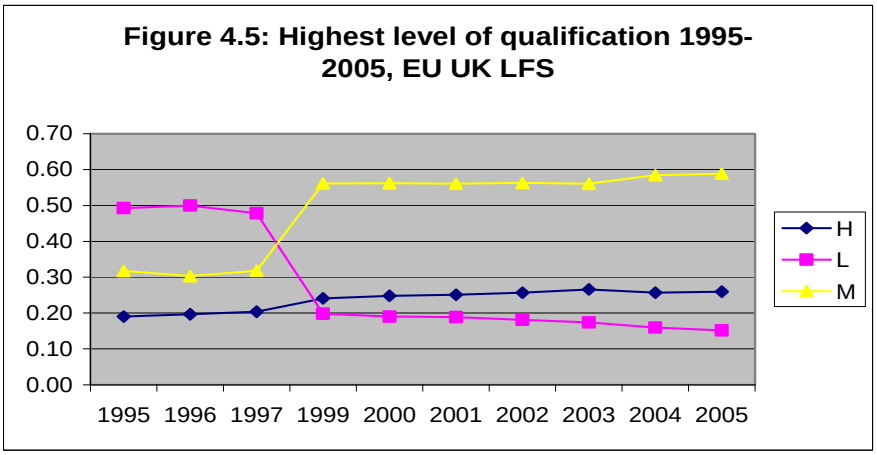
4.3 Implications for modelling and projecting the supply of qualifications

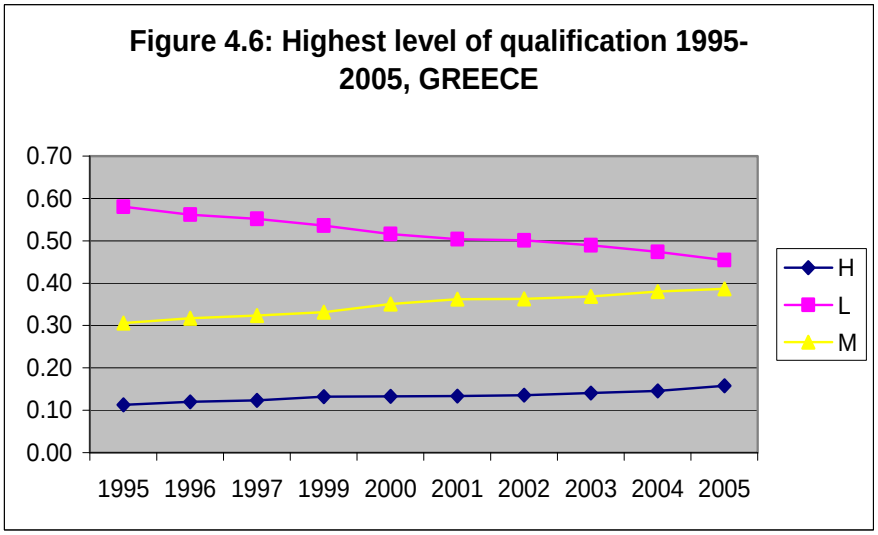
Qualifications are defined using ISCED (see Annex A for details). Regarding general trends in the three broad levels of education (the ISCED1D variable used in the LFS), a number of common and robust patterns can be observed. A negative trend is observed for the proportion with low qualifications in most countries. On the other hand, the trend for those with medium (M) and high (H) level qualifications is positive.

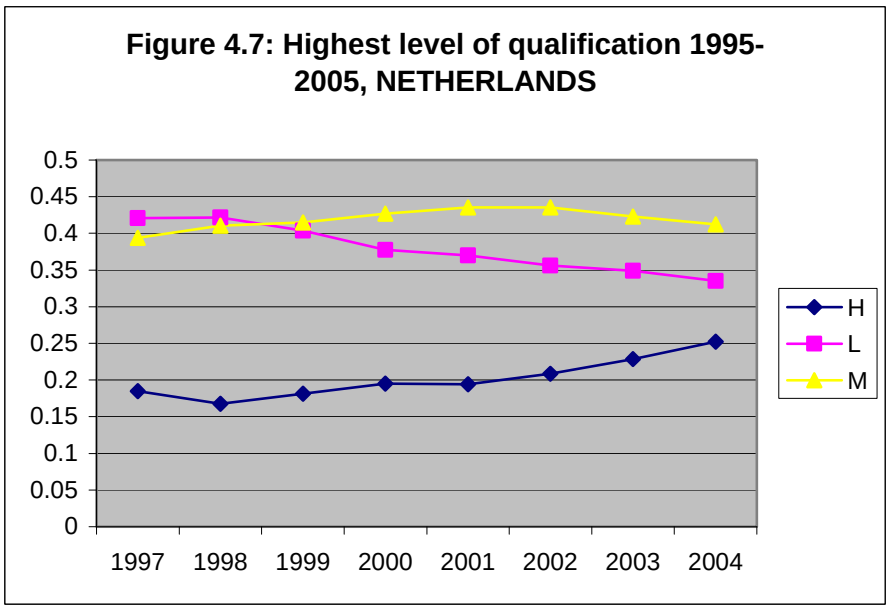
Even at this very broad level of aggregation some inconsistencies still exist over time. For instance, in the case of the UK (see Figure 4.5) there is a sudden jump between 1997/1998 in the proportion for these with medium (M) and low (L) levels of qualification. However, the data for Greece (Figure 4.6) are smoother and more consistent over time, no notable jumps are observed. Regarding the Netherlands (Figure 4.7), some kinks of small size are observed, and thus the trends are not as smooth as for Greece, but there are no obvious discontinuities.

However, even though the general patterns in these two countries are broadly smooth, the situation varies across the rest of the EU countries. For instance, for Denmark and Finland the patterns show a very erratic behaviour. A number of other problems with the data also exist. For instance, in Germany the EU LFS data are available for the very recent period from 2002. For Ireland there are no data for 1998.

The inconsistencies in the data are much more notable at a more detailed level of ISCED. So much so, that in the Cedefop project it was concluded that these problems were so severe that at present there is little merit in pursuing an analysis at the more detailed level of ISCED. The main focus of the Cedefop project on supply is therefore limited to the most aggregate level (high, medium, low qualifications).





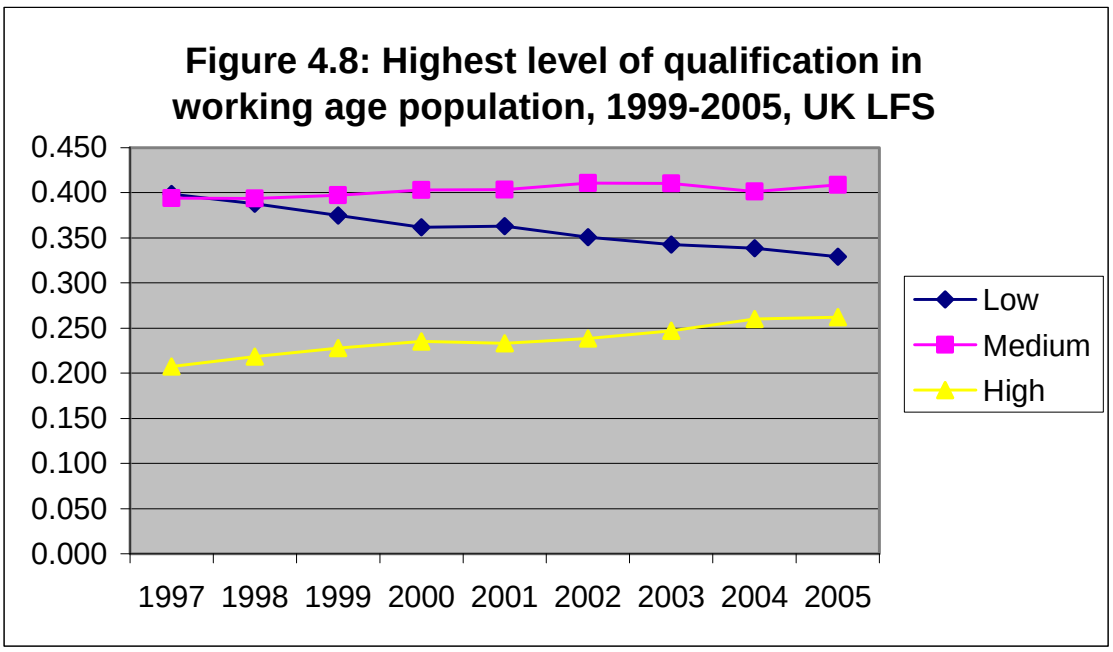


Source: IER estimates, based on Eurostata LFS data

It is also interesting to point out that the problems in the EU LFS data are not always reflected in the estimates from national LFS data. Figure 4.8 shows how the level of education is distributed for the working age population of the UK, based on the UK version of the LFS data. There are some very significant differences compared to those in Figure 4.5 based on the Eurostat version of the UK LFS data. Even using the simple classification into the three broad levels of education as recommended by Eurostat, there are significant differences from the results available from the national versions of these statistical surveys, even though they both use a common ISCED classification. This emphasises the need to clean up the Eurostat version of the dataset.

The results on the highest level of qualification held are sensitive to a range of issues, including precisely which vintage of LFS information is used, precisely how the data set is interrogated and the treatment of missing cases and "don't knows" (see Annex B for further discussion on this point). The UK based estimates suggest that the proportion of those with low qualifications is just over 30% whereas the estimates based on Eurostat data suggest a much lower figure of around 15%. This appears to be far too low and probably excludes those qualified at level 1.

Similar difficulties probably exist in other countries. In order to investigate how the levels of workforce educational attainment from the two sources differ from each other for each country, CEDEFOP has initiated a dialogue with experts from individual member states, who have been asked to provide CEDEFOP with their best estimates of national trends. These issues will be discussed in future Cedefop workshops.



Source: IER estimates, based on Eurostata LFS data

In the Cedefop project on the supply of qualifications a multinomial logistic regression model is used to estimate the probability of a representative individual attaining qualification level j at time t, based on an analysis of a combination of time series and cross-sectional data on individuals from the EU LFS micro dataset. This probability is expressed as a function of a vector of explanatory variables, normalised by the sum of probabilities for all qualification categories (for further details see Annex D). The main focus is on the 3 levels of qualifications as measured using ISCED. The sum of probabilities is constrained to add up to one. The vector of the regressors, which are included in the model as explanatory variables, include age group, gender, and time*gender and time*age group interactions. The regression coefficients are estimated so that the predicted model achieves 'best fit' to the observed data. This is done using the maximum likelihood method. The matrix of regression coefficients is then used to predict the distribution of people by qualification at each point in time, t. Categorical variables (age, gender, etc) are included in the model as an exhaustive set of dummy variables.

Some typical results obtained by applying the multinomial logistic regression technique are presented here. The results are for 3 countries: UK, Netherlands and Greece. Annex D presents more details on the econometric results. The results estimated are quite robust, and most of the coefficients were statistically significant. The robustness of the estimated models should therefore provide a reasonably sound foundation for obtaining predictions for the shares of educational qualifications attainment. However, the difficulties with the data outlined above do cause problems in some cases.

The coefficients from the model are used to project future educational attainment by unit of time, gender, and age-group. Figures 4.9, 4.10 and 4.11 show the results of this exercise. The Figures include both the historical and the future period. The bold line shows the fitted values of the historical data. In all cases, the fitted values match closely with the actual values, and the results for the future period seem plausible. In the case of the Netherlands, where the historical values show a more erratic pattern, the predicted future values show a very rapid rate of growth for those with high levels of qualifications.

While these results demonstrate the general feasibility of using the LFS to develop projections of the supply of people with different levels of qualifications, they also highlight the limitations of the data. An important general consideration when interpreting the results of all the econometric analysis is the relatively short period of historical data used in order to forecast the period ahead. In addition, it is important to emphasise that the methodology developed is based on the assumption that the future patterns will represent a continuation of those of the past, and does not take into consideration possible future changes in these patterns that might arise because of changed social and economic circumstances.

Figure 4.9: Qualifications Shares in working age population - UK

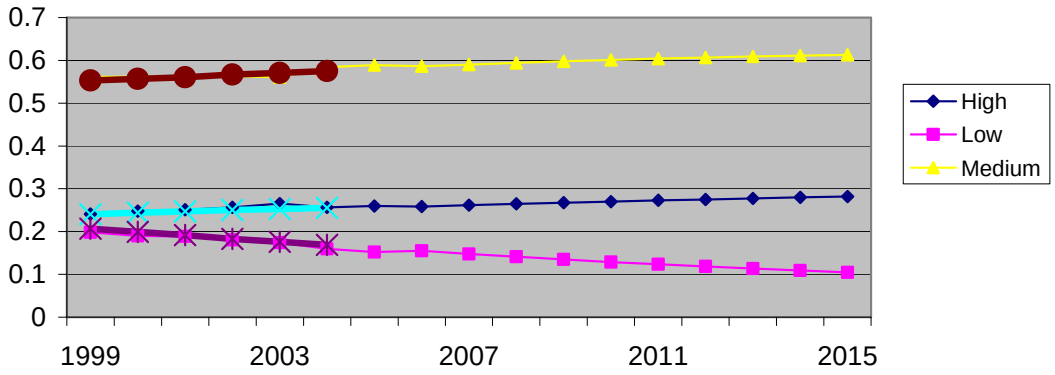


Figure 4.10: Qualifications Shares in working age population - GREECE

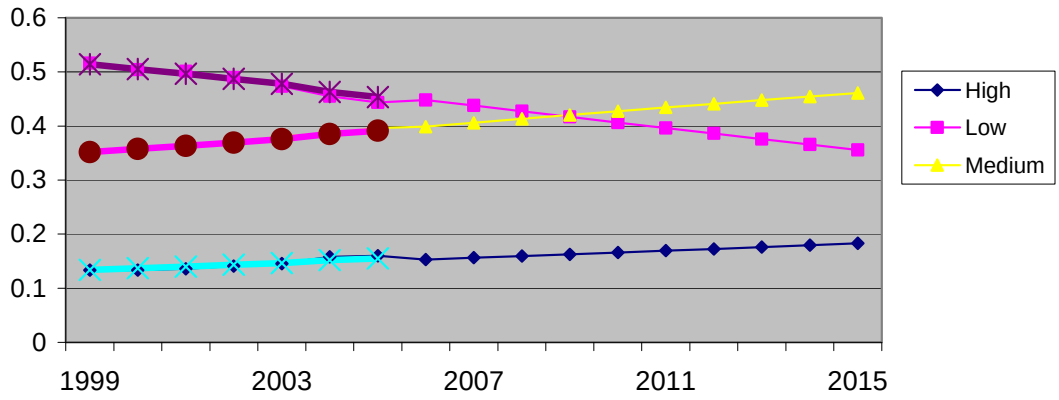
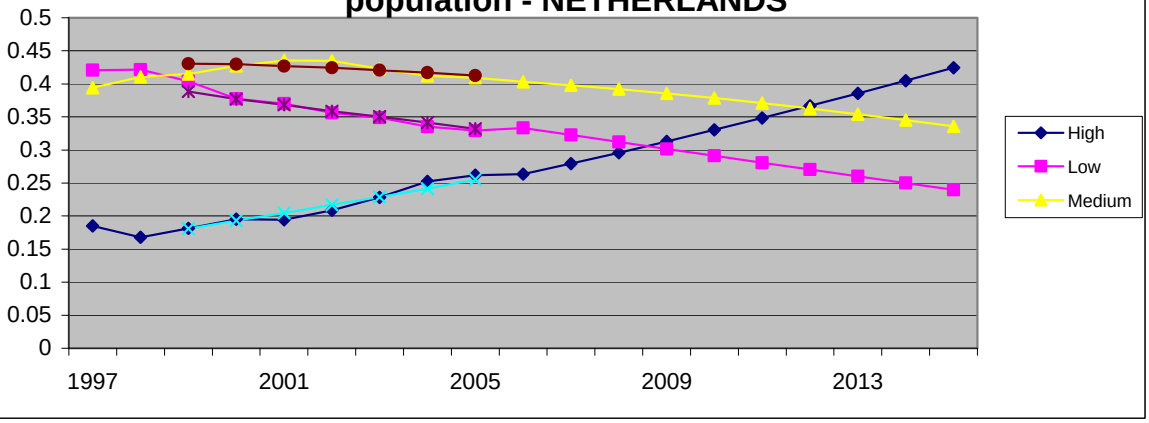


Figure 4.11: Qualifications Shares in working age population - NETHERLANDS



Source: IER estimates, based on Eurostata LFS data

5. Alternative ways to measure changing occupational employment structure

While it is clear that the LFS provides an important and valuable source of information on changing employment structure, it is also plain that it has a number of technical limitations, as set out above. Some, but by no means all, of these are a function of too small a sample size for identifying very detailed patterns within sectors.

Other limitations include the fact that the LFS only covers certain aspects of skill (notably occupation and qualifications). It does not deal with generic skills which are regarded as of increasing importance by many users.

Even if sample sizes could be increased there would still be problems of bias due to self and proxy responses. In general, the household based LFS is much better suited to dealing with issues relating to supply than demand for labour. For the latter it is important to get an employer's perspective.

Cedefop and others are currently considering whether a new Employer based survey is needed to monitor, and help to anticipate, changing skill needs. A Workshop on the use of employer surveys was held jointly with OECD in Paris, in May 2008.[9] This highlighted that there are potentially many disparate interests and possible foci for such survey at a pan-European level. These include surveys to monitor and measure:

i. changes in historical employment structure (the demand for skills);
ii. possible future skill needs;
iii. current skill shortages and skill gaps (including various job vacancy surveys);
iv. vacancy information based on public employment service administrative data;
v. recruitment practices;
vi. adaptation to change;
vii. links between skills and performance;
viii. continuing vocational training;
ix. different aspects of skills (such as key, core and generic skills);

One employer survey cannot meet all these needs. There is a need to focus on what is the key question to which answers are needed. What is most useful in terms of anticipating changing skill needs will depend on the on the different approaches to anticipation of skill needs that are adopted, as well as the state of the existing statistical infrastructure.

Skills can, of course, be measured various ways:

- occupation;
- qualification; and
- soft, key, core or generic skills.

[9] *Employers' surveys as a tool for identification of skill needs*. OECED/Cedefop Skillsnet Expert Workshop, 22-23 May 2008, Paris, OECD.

Scientific progress requires taxonomy and measurement for each of these. But the most fundamental gap across Europe is the first of these – especially robust information on occupation by sector. This is a crucial part of the statistical infrastructure. It is not just needed for projections. It is an essential element in understanding the current situation. Other things are also important but without these core data it is impossible to measure skill demand.

This is not just about taking employers views, but rather using their insights about what they actually do to monitor and measure actual skill demand. The focus is on real **needs**, what skills employers reveal they require (not necessarily what they perceive or say). How they behave is the key - who do they employ and in what positions? Other surveys can then help to translate this into demand for qualifications and soft skills.

The main method used across the world for projecting future skill needs, relies on quantitative analysis, based on formal econometric models (Wilson *et al*. 2004). However, good basic data on occupational employment structure are an essential prerequisite to building such models. There was some discussion in the Paris Workshop about the use of employer surveys as an **alternative** to such quantitative methods. If the prime objective is anticipating future skills needs, then it might appear that option ii. is the obvious choice. If we want to know what employers skill needs will be why not ask them? However, past experience suggests that employers are not very good at anticipating their future skill needs (see the review in Wilson, *et al*. (2004)). Such results have usually turned out to be biased, inconsistent and generally unreliable.

However, surveys of employers for many countries are a key element in helping to anticipate changing skill needs, not by asking employers questions about possible future skill needs, nor indeed about their perceptions of current skill deficiencies, but by focussing attention on what employers actually do by explicitly measuring their employment structures. Such surveys can provide crucial insights into current trends, as well as providing the basic data for building quantitative models that can help to anticipate future change. They focus on facts rather than opinions and perceptions.

There a number of examples of good and useful practice. In particular the Occupational Employment Statistics survey (OES) conducted by the Bureau of Labour Statistics in the USA stands out as delivering robust and very detailed data on both employment and pay that is highly valued by users. The OES started in 1968 and has gradually established a very sound statistical base upon to build a whole range of modelling and monitoring activities, including anticipation of changing occupational structure as carried out by the BLS. One of its prime supporters from a financial perspective is the agency responsible for monitoring migration flows. The survey provides crucial information on changing patterns of demand, as well as wages,

which can be used to help identify occupations for which there is a case to allow inward migration. The presence of data on pay (as well as numbers) provides the potential for analysing substitution effects, and developing growth accounting models.

The OES survey is very detailed including 800 plus occupations, which provides much greater insight into changing skill needs than the broad occupational categories possible using LFS data. The response rate is a very impressive 78%, which is very high for a voluntary survey. This is largely achieved through use of postal survey methods (backed up with internet feedback and query response) and emphasis on civic responsibility. The costs of conducting this survey, the related projection activities and the complementary O*NET system (which focuses in more detail on changing generic skill needs within occupations) are substantial.[10] This investment is supported by a government interested not in trying to plan the future but in informing its citizens so that they can make the decisions about what skills to invest in. While a number of individual countries within Europe have comparable data to that available from the OES, there is at present no pan-European equivalent.

The Paris workshop considered various possible options for developing employer based, pan-European surveys to fill this gap and to provide information that can help to improve the balance between the demand for and supply of skills, including better anticipation of changing skill needs across Europe. These included modifying a number of existing surveys, as well as developing new ones. Possible existing surveys include:

- The CVTS3 survey, which focuses on continuing vocational training;[11]
- In a number of countries there are existing employer based survey of current skill deficiencies (vacancies and skill gaps) which might be harmonised;
- Job vacancy data based on administrative sources (European Public Employment Services Vacancy Monitoring (EPVM));[12]
- The ongoing, OECD sponsored, PIACC initiative focusing on generic skills.[13]

While there was some heated debate about the various alternatives at the Paris Workshop it is clear that:

- Modifying the existing pan-European CVTS3 survey is not practicable. The need for a CVTS is essential, but this survey is focussing on different issues to those required if the main objective is to monitor and measure changing overall skill needs.

[10] For details of O*Net see: http://www.onetcenter.org/overview.html
The O*NET program is the US's primary source of occupational information. Central to the project is the O*NET database, containing information on hundreds of standardized and occupation-specific descriptors. The database, which is available to the public at no cost, is continually updated by surveying a broad range of workers from each occupation. Information from this database forms the heart of ***O*NET OnLine***, an interactive application for exploring and searching occupations. The database also provides the basis for Career Exploration Tools, a set of valuable assessment instruments for workers and students looking to find or change careers.
[11] European enterprise survey focused on continuing vocational training (CVT) in enterprises.
[12] See Di Domenico, (2008).
[13] Programme for the International Assessment of Adult Competencies (PIACC).

- Adjusting / harmonising existing national surveys focussed on “shortages” or “gaps” is also unrealistic. They are too disparate, and there are too many vested interests which would resist changing the current national formats, to make this a practicable proposition. Such surveys can provide some complementary information but they remain focussed on the margin of the labour market (the current difference between supply and demand), which is inevitably ephemeral as markets adjust to deal with any immediate problems.
- Similar remarks apply to the EPVM. It is unlikely that such administrative data systems can be modified to cover labour demand as opposed to vacancies. Such information provides a useful complement to data on employment levels, but US and other experience suggests that such information is not an essential component in measuring changing skill needs. Vacancies focus on the margins of the labour market and tend to be very ephemeral unless there are long-standing market failures.
- The PIACC initiative should go a long way to filling the data gap on generic or soft skills. There is also the possibility of exploiting the work the US has done in O*NET, on the grounds that for many occupations the details of skills involved will be common across the Atlantic. However there seems little prospect of combining the PIACC survey with a more basic survey to measure occupational structure at as detailed level;
- Other surveys focussing on recruitment practices, adaptation to change and links between skills and performance can also contribute some very interesting insights, but they tend to require much more complex survey designs. It is arguable that other research will focus in much greater detail on improving understanding of how organisations and labour markets work. Such research does not necessarily require a pan-European dimension. The main gap at a pan-European level is for much more basic information about the current patterns of skill demand.

There seems to be a strong *prima facie* case for a new employer based, pan-European survey of employers, but there remains considerable debate about what its focus should be.

The biggest gap from the evidence presented in this paper is robust information on the current patterns of skill demand by occupation within sectors. The LFS data available at European level has been taken as far as it can be in the Cedefop project. To go a stage further will require new a better data. Increasing the LFS sample size can help to resolve some of the problems identified above but it cannot solve the basic difficulties of imprecision and bias as a result of self and proxy reporting in a household based survey. An employer focussed survey is essential to get a proper demand perspective.

The US is some 40 years ahead of Europe with its OES, but it is possible to catch up quite quickly if decisions are made promptly. However a serious investment is needed. Cedefop can play a key role by funding a study to explore these issues in greater depth and to pilot alternative approaches that could be tried at a pan-European level. Of course the OES is not perfect but it, and the history of its development, has some valuable lessons for Europe.

6. Main Implications and Concluding Remarks

6.1 The LFS does facilitate analysis of prospects for both labour demand and supply

Data available to conduct a pan-European analysis of changing employment structure by occupation and qualification have improved significantly in Europe over the past few years, with the publication of harmonised statistics from the LFS. However, the analysis in this paper suggests that there is still some way to go before these data can be regarded as robust and harmonised enough to develop sophisticated economic models. Very short time series and lack of consistency combine to make anything other than fairly simple methods of modelling occupation and qualification shares impractical at the kind of detailed level attempted here.

At the broadest level, the main trends in occupational and qualification patterns are reasonably clear. Delving down into greater detail, the quality of the data often obscures any patterns. Much of this is probably statistical "noise" which tends to cancel out as data are aggregated. This is of less comfort if one is interested in the detail for its own sake, both in terms of understanding structural changes in economy (the industry dimension), and changes in the way work is carried out (the occupational dimension and the demand for skills). Although there may be some scope for improving the sophistication of the modelling work if industries and occupations were aggregated together, this would be at the expense of this desired emphasis on detail. The latter is essential to provide both policy makers and individuals with the information they need to make sensible decisions about investment in skills.

The projects undertaken for Cedefop indicate that, using simple methods and some judgement, it is possible to develop some reasonably robust results on likely changes in employment patterns at a pan-European level. Nevertheless, it is clear that there are still many problems with the data for individual countries. This needs further input from individual country experts and the relevant statistical authorities to resolve.

Until the quality and length of the time series data available have both been increased the potential benefits of more sophisticated modelling work are probably limited. There is probably some potential to do this on a country by country basis by using national specific data. The benefits in terms of increased precision and statistical robustness would be considerable. However, this would move away from the advantage of using a common data set, which facilitates comparison across countries.

6.2 Problems with the LFS and some possible solutions

The main problems identified with the LFS data relate to the quality of the most detailed information by sector, occupation and qualification. There are particular problems of missing data for some cells in the industry by occupation by qualification arrays, as well as difficulties

caused by changing systems of classification (resulting in breaks in time series), and the lack of complete harmonisation across countries. In combination, these problems can result in gaps or very short time series of consistent estimates available for analysis. Even where estimates do exist there are often concerns about precision and robustness. Work to improve the quality of these data should be a priority for Eurostat and national statistical authorities.

From the perspective of improving the work on anticipating changing skill needs, the main emphasis should be on improving the quality of the industry by occupation (and qualification) estimates that underlie the Cedefop work. This paper has explored this issue in some detail, highlighting particular areas of concern and making recommendations about where improvements might be possible. These include:

1. Improvement of existing LFS data by Eurostat;
2. Improvement of existing LFS data by Member States;
3. Extension of the existing LFS data set (i.e. increasing sample size);
4. Enhancement of the existing LFS data set (i.e adding new questions or refining old ones, to generate more useful explanatory variables);

With regard to improvements to the existing data set, this should include:

5. Harmonisation issues across countries (involving both Eurostat and Statistical authorities from Member States). Problems include:
    - Occupations (despite the use of ISCO there are still significant differences);
    - Qualifications (lack of clarity with regard to treatment of apprenticeships, and various other qualifications).
6. Problems of consistency over time (again involving both parties), especially to deal with known changes in national systems of classification; and finally,
7. There is a strong case for the development of a new employer based survey to complement the LFS household survey.

Section 5 of this paper has explored the case for a new employer based survey to measure changing demand for skills. Only some of the weaknesses of the LFS can be dealt with by a further boost to sample size and by improvements in harmonisation and avoidance of discontinuities over time. Others require a more fundamental re-examination of what it is that needs to be measured and what is the best way of doing it (both conceptually as well as in terms of value for money). An employer survey which focuses specifically on employers' revealed demand, focussing on what occupations they actually employ, is arguably the best way to do this. It avoids problems of bias and lack of understanding of occupational definitions. There is an urgent need for strategic investment in such data at a European level to deal with these concerns if the objectives for better anticipation of changing skill needs set out in the *Employment Guidelines* are to be achieved.

References

Becker, G. (1964, 1993, 3rd ed.). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. Chicago, University of Chicago Press.

BLS (1997) *BLS Handbook of Methods*, Chapter 13, Employment Projections, Acemogly, D. (2002) Technical Change, Inequality, and the Labour Market, Journal of Economic Literature, 40,1, pp.7-72

Briscoe, G. and R. Wilson (2003) Modelling UK occupational employment, *International Journal of Manpower*, 24,5, pp. 256-289.

Collings R. (1979) Credential Society: A Historical Sociology of Education and Stratification, Academic Press, ISBN-10: 0121813606

Di Domenico, Germana (2008) *Job Vacancies and Recruitment Difficulties. Which observation instruments are available at EC level and how to improve them?*, Employers' surveys as a tool for identification of skill needs. Skillsnet Expert Workshop organised by Cedefop and OECD, 22-23 May 2008, Paris

European Communities (2003) *The European Union Labour Force Survey: Methods and definitions 2001*, European Commission. Luxemburg.

Eurostat (2007) Comparability Of Results, Breaks In Series And Coherence With Other Statistics.
http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/LFS/LFS_COMPARABILITY.htm

Fox R. and B. Comerford (2006), Cohort component method of deriving replacement demand, Paper presented at the workshop Anticipating Europe's Skill Needs, 2-3 November 2006, University of Warwick, UK

Elias, Peter and Margaret Birch (2008) Proposals for the redefinition and classification of managerial occupations in the 2010 revision of the Standard Occupational Classification, paper submitted to ONS committee responsible for revising SOC, Institute for Employment Research, University of Warwick

Hecher, D. (2005) Occupational Employment Projections to 2004-2014, Monthly Labor Review, 2005, pp. 70-102.

Mark Keese, (2008) Progress with the JRA pilot. Joint OECD-Cedefop Workshop on Employers' Surveys as a Tool for Identification of Skill Needs, Paris, 22 – 23 May 2008.

Lauer, C. (2006) A model of educational attainment, Application to the German Case, Discussion Paper No. 02-06, ZEW, Centre for European Economic Research.

Layard, R. and G. Psacharopoulos (1974) The screening hypothesis and the returns to education The Journal of Political Economy, Vol. 82, 5, pp. 985-998

Livanos, I and R. Wilson (2008) *Trends in stocks of those acquiring qualifications: Multilogit models based on LFS data.* paper presented at the Workshop on Medium-term forecast of skills supply in Europe: interim results, Skillsnet Technical Workshop 2-3 June 2008, Thessaloniki, Greece.

Livanos, I. and R. Wilson (2007a) *National Accounts vs. Labour Force Survey: exploring differences in concepts, definitions, and numbers of employment*, Medium-term forecasts of occupational skill needs in Europe: interim results, 07-08, May 2007, Maastricht, the Netherlands.

Livanos, I. and R. A. Wilson, (2007b). *Expansion Demand and Changes in Occupational and Qualifications Structure.* paper presented at the Workshop on Medium-term forecast of occupational skill needs in Europe: Final results. 4th TECHNICAL WORKSHOP, November 8-9, Vienna.

Stimpson, Alex , (2008) CVTS3: an update. OECD/Cedefop Skillsnet expert workshop: Employers' surveys as a tool for identification of skill needs, 22-23 May 2008, , OECD, Paris.Wilson, R.A. (2007), *Trends in Employment Creation in Europe*. paper prepared for the Portuguese Presidency Conference on Employment In Europe - Prospects and Priorities, Lisbon, 8-9 October 2007

Wilson, R.A. *et al.* (2008) *Future skill needs in Europe Medium-term forecast: Synthesis Report*, Cedefop / Luxembourg: Office for Official Publications of the European Communities, ISBN 978-92-896-0500-7

Wilson, R.A. and I. Livanos (2007) *Expansion Demand and Changes in Occupational Structure*, Paper presented at the Workshop on Medium-term forecast of occupational skill needs in Europe: Interim results. 3rd Technical Workshop, May 7-8, Maastricht, Netherlands.

Wilson, R. A., I. Woolard and D. Lee (2004), *Developing a National Skills Forecasting Tool for South Africa*, South African Department of Labour: Pretoria. (50 pages plus extensive Annexes) (see especially Annex B: Labour Market Projections: A Review of International Best Practice).

Acronyms

Institutions and organisations

BLS	Bureau of Labor Statistics (US Department of Labor)
CE	Cambridge Econometrics
Cedefop	European Centre for the Development of Vocational Training
IER	Institute for Employment Research, University of Warwick
ROA	Research Centre for Education and the Labour Market, University of Maastricht
Skillsnet	Cedefop network on Early Identification of Skill Needs

Others

CVTS	Continuing Vocational Training Survey
EU	European Union
EU25	European Union members as of December 2006
Eur27	EU25 plus Norway and Switzerland
EPVM	European Public Employment Service Vacancy Monitoring
LFS	Labour Force Survey
NA	National Accounts
ISCO	International Standard Classification of Occupation
ISCED	International Standard Classification of Education
NACE	Statistical Classification of Economic Activities in the European Community
OES	Occupational Employment Statistics survey (USA)
O*NET	US system for organising occupational information
PIACC	Programme for the International Assessment of Adult Competencies
U-O-E	UNESCO-OECD-Eurostat

ANNEX A: Defining and measuring Qualifications ISCED

Educational Classification ISCED

ISCED 0 — PRE-PRIMARY EDUCATION

Programs at level 0, (pre-primary) defined as the initial stage of organised instruction, are designed primarily to introduce very young children to a school-type environment, i.e. to provide a bridge between the home and a school based atmosphere. Upon completion of these programs, children continue their education at level 1 (primary education).

ISCED 1 — PRIMARY EDUCATION OR FIRST STAGE OF BASIC EDUCATION

Programmes at level 1 are normally designed on a unit or project basis to give students a sound basic education in reading, writing and mathematics along with an elementary understanding of other subjects such as history, geography, natural science, social science, art and music. In some cases religious instruction is featured. The core at this level consists of education provided for children, the customary or legal age of entrance being not younger than five years or older than seven years. This level covers, in principle, six years of full-time schooling.

ISCED 2 — LOWER SECONDARY EDUCATION OR SECOND STAGE OF BASIC EDUCATION

The contents of education at this stage are typically designed to complete the provision of basic education which began at ISCED level 1. In many, if not most countries, the educational aim is to lay the foundation for lifelong learning and human development. The programmes at this level are usually on a more subject oriented pattern using more specialised teachers and more often several teachers are conducting classes in their field of specialisation. The full implementation of basic skills occurs at this level. The end of this level often coincides with the end of compulsory schooling where it exists.

ISCED 3 — (UPPER) SECONDARY EDUCATION

This level of education typically begins at the end of full-time compulsory education for those countries that have a system of compulsory education. More specialisation may be observed at this level than at ISCED level 2 and often teachers need to be more qualified or specialised than for ISCED level 2. The entrance age to this level is typically 15 to 16 years. The educational programmes included at this level typically require the completion of some 9 years of full-time education (since the beginning of level 1) for admission or a combination of education and vocational or technical experience.

ISCED 3A: Programmes designed to provide direct access to ISCED 5A;

ISCED 3B: Programmes designed to provide direct access to ISCED 5B;

ISCED 3C: Programmes not designed to lead to ISCED 5A or 5B.

ISCED 4 POST-SECONDARY NON TERTIARY EDUCATION

ISCED 4 captures programmes that straddle the boundary between upper secondary and post-secondary education from an international point of view, even though they might clearly be considered as upper secondary or post-secondary programmes in a national context. These programmes can, considering their content, not be regarded as tertiary programmes. They are often not significantly more advanced than programmes at ISCED 3 but they serve to broaden the knowledge of participants who have already completed a programme at level 3. Typical examples are programmes designed to prepare students for studies at level 5 who, although having completed

ISCED level 3 did not follow a curriculum which would allow entry to level 5, i.e. pre-degree foundation courses or short vocational programmes. Second cycle programmes can be included as well.

ISCED 4A: See text for ISCED 3

ISCED 4B: See text for ISCED 3

ISCED 4C: See text for ISCED 3

ISCED 5 — FIRST STAGE OF TERTIARY EDUCATION (NOT LEADING DIRECTLY TO AN ADVANCED RESEARCH QUALIFICATION)

This level consists of tertiary programmes having an educational content more advanced than those offered at levels 3 and 4. Entry to these programmes normally requires the successful completion of ISCED level 3A or 3B or a similar qualification at ISCED level 4A. They do not lead to the award of an advanced research qualification (ISCED 6). These programmes must have a cumulative duration of at least two years.

ISCED 5A: Programmes that are largely theoretically based and are intended to provide sufficient qualifications for gaining entry into advanced research programmes and professions with high skills requirements.

ISCED 5B: Programmes that are practically oriented/ occupationally specific and are mainly designed for participants to acquire the practical skills and know-how needed for employment in a particular occupation or trade or class of occupations or trades, the successful completion of which usually provides the participants with a labour-market relevant qualification

ISCED 6 — SECOND STAGE OF TERTIARY EDUCATION (LEADING TO AN ADVANCED RESEARCH QUALIFICATION)

This level is reserved for tertiary programmes which lead to the award of an advanced research qualification. The programmes are therefore devoted to advanced study and original research

and not based on coursework only. They typically require the submission of a thesis or dissertation of publishable quality which is the product of original research and represents a significant contribution to knowledge. They prepare graduates for faculty posts in institutions offering ISCED 5A programmes, as well as research posts in government, industry, etc.

ISCED classification and the variables used in the LFS

Educational attainment in the Eurostat LFS is classified by the International Standard Classification of Education (ISCED) into 19 detailed categories associated with different levels of education. These are referred to as ISCED2D in the LFS. Based on the ISCED2D variable, information can be aggregated into the 7 main levels as shown in Table A.1 or into 3 much broader ISCED categories as shown in Table A.2. The latter is known in the LFS as ISCED1D.

The more detailed ISCED2D variable is not necessarily reliable since, according to the *EU LFS Users Guide* (2003), it is available from 1998 only and often lacks comparability across countries. The variable ISCED1D was therefore chosen as the prime focus of analysis in the Cedefop project.

Table A.1: ISCED – The 7 main levels of education

Level 0:	Pre-Primary Education;
Level 1:	Primary Education or First Stage of Basic Education;
Level 2:	Lower Secondary or Second Stage of Basic Education;
Level 3:	Upper Secondary Education;
Level 4:	Post - Secondary, Non-Tertiary Education;
Level 5:	First Stage of Tertiary Education (not leading directly to research qualification);
Level 6:	Second Stage of Tertiary Education (leading to research qualification)

ISCED1D distinguishes three broad levels of education: high, medium and low. In particular, ISCED1D combines ISCED's seven levels of education as shown in Table A.2.

Table A.2: LFS variable ISCED1D - 3 levels of education

Low:	ISCED 0-2
Medium:	ISCED 3-4
High:	ISCED 5-6

ISCED2D is a more detailed LFS variable that distinguishes 19 levels of education, derived from the two-digit categories in ISCED. These 19 levels are described in Table A.3.

Table A.3: LFS variable ISCED2D - 19 levels of education

No formal education or below ISCED 1
ISCED 0-1
ISCED 1
ISCED 2
ISCED2D
ISCED 3c (shorter than 3 years)
ISCED 3 (without distinction a,b or c possible, 3 y+)
ISCED 3c (3 years or more)
ISCED 3a,b
ISCED 3c (3 years or more) or ISCED 4c
ISCED 3b or ISCED 4b
ISCED 3a or ISCED 4a
ISCED 3 or 4 (without distinction a,b or c possible)
ISCED 4a ,b
ISCED 4c
ISCED 4 (without distinction a,b or c possible)
ISCED 5b
ISCED 5a
ISCED 6

ANNEX B: Estimating Qualifications Patterns using LFS data

General issues

Estimating the proportion of the population qualified at different levels using LFS data is not as straightforward a task as it might at first appear. Results can differ depending on a number of factors, including:

1. The vintage of the data set (in the UK, for example, the Office for National Statistics (ONS) release different versions, incorporating corrections and adjustments based on other data to benchmark the numbers more robustly);
2. There are also different versions of the LFS available at the same time which can contain different information. (for example, UK Government Departments get uncensored access to the data, while public versions of the data set have various information suppressed because of concerns about confidentiality and data protection issues - Eurostat get a different version again);
3. The coverage over time (the LFS is conducted quarterly, using an overlapping sample and there are some differences in the information collected in each quarter. Quarters can be combined together to create an annual average but this require careful treatment to avoid double counting of some individuals due to the same individual being questioned in subsequent quarters);
4. There are often a number of slightly different questions relating to qualifications (resulting in a choice of possible variables in the final survey results);
5. The allocation of individual qualifications to ISCED levels is often not straightforward (in the UK for example, the mapping is complicated by the need to recognise that in some cases only a proportion of individuals have achieved the threshold levels to move them up from one NQF level to the next (depending on grades achieved, etc).[14] This information is not available in the LFS and an apportionment based on other information is needed. This requires a procedure to randomly select and allocate individual cases which can lead to sampling variation of estimates taken from the LFS, separate from the normal statistical variation.
6. The LFS is a complex survey and routing through the questionnaire can affect the number of missing cases. The proportions with different qualifications can be affected by this. Unless this routing is dealt with in precisely the same way when interrogating the survey, slightly different results can be obtained.

[14] NQF is the UK National Qualification Framework (levels of which are broadly consistent with ISCDED).

7. However the issue of routing is dealt with, there are always a significant number of cases missing. In addition, in a significant number of cases the individual responds that they “Don’t know” the qualifications they hold. These types of problems can be dealt with in various different ways. One possibility is to assume that such cases all have the same probabilities of holding qualifications as the population as a whole. Another is to assume that those who don't know have no qualifications. A third possibility is to allocate them all to a residual category. This can lead to quite different outcomes.
8. Variations can also arise because of the use of a different population (for example excluding all those above the official retirement age).
9. Differences can also arise because of the focus sometimes on the highest qualification held as opposed to all qualifications held.

Unless all of these factors are common then two independent interrogations of the LFS can lead to very different outcomes. Without very detailed documentation on how data were extracted and estimates made it is often not possible to exactly replicate results. Thus it is no surprise that the results in the national and Eurostat versions of the data can often differ.

ANNEX C: LFS versus national Accounts estimates of employment

When the aim is to focus on developments of occupational and qualification patterns within industries (as is required to link into a detailed macroeconomic models such as that used in the Cedefop project), then a number of problems emerge relating to the comparability between the LFS and national accounts based estimates. This issue is discussed in Box 2.1 taken from Wilson *et al.* (2008).

Box C.1: Relation between employment in the labour force survey and in national accounts

Estimates of employment in national accounts may differ from results of other statistics and surveys, in particular the labour force survey (LFS). There are differences due to integration of sources and for conceptual reasons:

Differences due to integration:

- National accounts integrate information from many sources. All sources available (including LFS) are assessed and subsequently the best way of integrating them is decided. Each source may shed light on a part of the economy. Some countries make very minor use of LFS in national accounts. The information is combined to provide the most complete and consistent estimate. As a consequence, each individual basic source may provide results that are different from the integrated national accounts estimates.
- In national accounts, employment figures must be consistent with other variables such as output and compensation of employees (i.e. wages, salaries and social contributions). Ensuring consistency between variables may result in adjustments.

Conceptual differences:

- Geographical scope: ESA95 acknowledges two employment concepts depending on the geographical coverage: resident persons in employment (i.e. the so-called national concept of employment) and employment in resident production units irrespective of the place of residence of the employed person (i.e. domestic concept). The difference between them corresponds mainly to the net number of cross-border workers. The figures in this News Release correspond to the domestic concept. This concept is more appropriate when examining employment and GDP together. LFS, on the other hand, covers resident households. Hence LFS gives information on the major part of the national concept. This means that
- LFS data must be adjusted, mainly for cross-border workers, to align with the domestic concept normally used in national accounts.
- Coverage differences: 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. They are all covered by ESA95 employment. Appropriate adjustments are therefore needed.
- Recording thresholds: LFS results exclude persons below 15 years old from the definition of employment (in some countries the exclusion boundaries are below 16 years old and/or above 75 years old). National accounts do not exclude individuals from employment because of age. The difference is very small in developed economies.

The size of these conceptual adjustments is modest, with the possible exception of conscripts, and cross-border workers for small countries.

Source: Eurostat 2007: http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/LFS/LFS_COMPARABILITY.htm

ANNEX D: Mlogit methods

D.1 Modelling occupational employment shares

A multinomial logistic regression model can be used to estimate the probability of an individual in particular industry working in occupation (OCC) *j* at time *t*. The general model is specified as follows:

$$\Pr\left(OCC = j \mid T = t\right) = \frac{\exp\left(\Omega^{(j)} X\right)}{1 + \sum_{i=1}^{N} \exp\left(\Omega^{(i)} X\right)} \quad (1)$$

and for an arbitrarily chosen base category:

$$\Pr\left(OCC = N \mid T = t\right) = \frac{1}{1 + \sum_{i=1}^{N} \exp\left(\Omega^{(i)} X\right)} \quad (2)$$

The equations state that the probability of the representative individual working in a particular industry in occupation *j* at time *t* can be expressed as a function of explanatory variables, normalised by the sum of probabilities for all categories. There are N occupations and the sum of probabilities is constrained to add up to one (i.e. shares of employment across occupations sum to 100 percent).

Similarly, the above model can be used in order to estimate the probability of an individual working in occupation (OCC) *j* at time *t* holding a qualification *n*.

In the model **X** relates to a vector of regressors which are included in the model as explanatory variables for occupational or qualification structure. The time trend variable proxies the impact of technology and other factors on changing occupational structure. Time was interacted with country and industry to allow for these effects to be country and sector specific. The regression coefficients are estimated so that the predicted model achieves 'best fit' to the observed data.

This is done using the maximum likelihood method. In the model, Ω is the matrix of regression coefficients which is used to predict the distribution of employment by occupation at each point in time, *t*. Categorical variables (SIC, country, age, gender, etc) are included in the model as an exhaustive set of dummy variables. The primary aim of estimating such a model is to provide a behavioural explanation of changing patterns of employment structure. Such a model can be used to forecast shares of employment.

Table D.1 provides an illustration of the econometric output regarding the modelling of occupational shares for one particular case: *Printing & Publishing* in Greece. The table reports the coefficients, the Z-ratios and the P-values. An important feature of the multinomial logit

model is that it estimates *k-1* models, where *k* is the number of categories distinguished in the dependent variable. In this case there are 27-1=26 occupations. In the table *precision, handicraft, craft printing and related trades workers* are the base outcome. Therefore the estimated coefficients of the model for any other occupational group are relative to *precision, handicraft, craft printing and related trades workers*.

Since the parameter estimates are relative to the reference group, the standard interpretation of the multinomial logit is that for a unit change in the predictor variable, the logit of outcome *m* relative to the referent group is expected to change by its respective parameter estimate. For example, the multinomial logit for *Legislators and senior officials* for industry *Printing & Publishing* suggests that for one unit of change in time, the probability of being in the occupational group of Legislators and senior officials is expected to change by 0.010 relative to *precision, handicraft, craft printing and related trades workers* (reference category).

D.2 Econometric specification for the supply modelling

The multi-logit specification is also used to model the propensity of a representative individual in each country to obtain a given level of qualification, based on an analysis of a combination of time series and cross-sectional data on individuals from the EU LFS micro dataset. The same basic model can be applied to the total population, and the economically active population. This probability is expressed as a function of a vector of explanatory variables, normalised by the sum of probabilities for all qualification categories. The main focus is on the 3 levels of qualifications as measured using ISCED. The sum of probabilities is constrained to add up to one. The vector of the regressors, which are included in the model as explanatory variables, include age group, gender, and time*gender and time*age group interactions. The regression coefficients are estimated so that the predicted model achieves 'best fit' to the observed data. This is done using the maximum likelihood method. The econometric specification can be summarized as follows:

The probability of an individual holding a particular level of qualification (Q) *j* at time *t* is:

$$\Pr\left(Q = j \mid T = t\right) = \frac{\exp\left(e^{(j)} X\right)}{1 + \sum_{i=1}^{N} \exp\left(e^{(i)} X\right)} \qquad (3)$$

and for an arbitrarily chosen base category:

$$\Pr\left(Q = N \mid T = t\right) = \frac{1}{1 + \sum_{i=1}^{N} \exp\left(e^{(i)} X\right)} \qquad (4)$$

Table D.1: Greece multinomial logistic regression output

Printing & Publishing	Coefficient	Z-ratio	P-value
Legislators and senior officials	0.010	0.99	0.32
Corporate managers	-0.028	-24.94	0.00
Managers of small enterprises	0.028	92.91	0.00
Physical, mathematical and engineering science professionals	0.014	6.25	0.00
Life science and health professionals	-0.051	-5.61	0.00
Teaching professionals	-0.033	-3.72	0.00
Other professionals	0.005	11.85	0.00
Physical and engineering science associate professionals	0.047	69.36	0.00
Life science and health associate professionals	-0.033	-3.72	0.00
Teaching associate professionals	0.010	0.99	0.32
Other associate professionals	0.047	73.31	0.00
Office clerks	-0.003	-6.07	0.00
Customer services clerks	0.071	42.18	0.00
Personal and protective services workers	-0.113	-18.94	0.00
Models, salespersons and demonstrators	0.065	43.11	0.00
Skilled agricultural and fishery workers	-0.021	-2.19	0.029
Extraction and building trades workers	-0.135	-28.4	0.00
Metal, machinery and related trades workers	-0.087	-29.07	0.00
Other craft and related trades workers	-0.073	-13.92	0.00
Stationary plant and related operators	-0.093	-9.8	0.00
Machine operators and assemblers	-0.032	-37.02	0.00
Drivers and mobile plant operators	0.024	22.11	0.00
Sales and services elementary occupations	-0.053	-29.92	0.00
Agricultural, fishery and related labourers	0.010	0.99	0.32
Labourers in mining, construction, manufacturing and transport	-0.040	-22.15	0.00

Note: (occupation == Precision, handicraft, craft printing and related trades workers is the base outcome)

The equations state that the probability of the representative holding a qualification *j* at time *t* can be expressed as a function of explanatory variables (X), normalised by the sum of probabilities for all categories. There are N qualifications and the sum of probabilities is constrained to add up to one (i.e. shares of employment across qualifications categories sum to 100 percent).

Some typical results obtained by applying the multinomial logistic regression technique are presented here. The estimated model uses the specification outlined in equation 4. Table D.2 shows a sample of the econometric results, refering to the working age population (15-65) of the UK (using only data from 1998 onwards). The full results cover all other countries and also the economically active labour force as well as the population of working age. Table D.2 reports the coefficients for high (column 1) and low (column 2) levels of qualification; the medium level was set as the reference category. The standard errors are reported in parenthesis. An important feature of the multinomial logit model is that it estimates *k*-1 models, where *k* is the number of categories distinguished in the dependent variable. In this case there are 3-1=2 levels of education. The estimated coefficients of the model for any other qualification category are relative to the base outcome[15]. The last column of Table D.2 shows the means of the variables in the sample.

Since the parameter estimates are relative to the reference group, the standard interpretation of the multinomial logit is that for a unit change in the predictor variable, the logit of an outcome *m* (which in this case is either high or low level of qualification) relative to the referent group is expected to change by its respective parameter estimate. For example, the multinomial logit for the variable FEMALE for high level of education suggests that the log of the ration of the two probabilities P(HighQual)/P(MediumQual) is -16.75. This suggests that a female in the UK (at the time that the survey was conducted) had 16% lower probability of holding a high level than a medium level, of qualification compared to the reference category (male). Similarly, females have 108% higher probability of having a low level of education than a medium level of education. As for the time trended variables, the 0.00836 coefficient of the TIME*FEMALE variable suggest that for one unit of change in Time, the probability of a female holding a high qualification will increase by 0.008.

The coefficients obtained are then used in order to project future educational attainment by unit of time, gender, and age-group.

[15] This is chosen automatically by the Stata package and therefore differs from country to country.

Table D.2: Typical results for the UK, working age population

	(1)	(2)	
COEFFICIENTS	HIGH QUAL.	LOW QUAL.	MEAN
FEMALE	-16.75***	108.3***	51.7
	(5.858)	(6.793)	
TIME*FEMALE	0.00836***	-0.0539***	18.2
	(0.00293)	(0.00339)	
14-24 AGEGROUP	34.22***	-52.67***	21.4
	(11.20)	(10.09)	
25-34 AGEGROUP	-85.49***	-27.02**	
	(7.193)	(10.82)	
45-54 AGEGROUP	47.87***	123.3***	20.7
	(7.566)	(8.497)	
55-64 AGEGROUP	41.07***	163.6***	16.7
	(9.529)	(8.828)	
TIME * 14-24 AGEGROUP	-0.0177***	0.0262***	
	(0.00559)	(0.00504)	
TIME * 25-34 AGEGROUP	0.0428***	0.0133**	
	(0.00359)	(0.00541)	
TIME * 45-54 AGEGROUP	-0.0239***	-0.0613***	
	(0.00378)	(0.00424)	
TIME * 55-64 AGEGROUP	-0.0205***	-0.0812***	
	(0.00476)	(0.00441)	
Constant	-0.670***	-1.601***	
	(0.00717)	(0.0094	

Notes:

Observations 526,776

R-squared 0.0365

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Medium qualification is reference category