Using LFS Data for Cross-National Research: Promises, Examples, and Problems

Walter Müller and Markus Gangl  
MZES, University of Mannheim  
P.O. Box, D-68131 Mannheim, Germany  
e-mail wmueller@sowi.uni-mannheim.de or Markus.Gangl@mzes.uni-mannheim.de

Presentation to the  
International Workshop on Comparative Data on Education-to-Work Transitions  
OECD, Paris, 21-23 June 2000

The European Labour Force Survey (EULFS) is undoubtedly one of the core data bases for the comparative study of European Societies. By its very nature it is a particularly reach data base relating to various conditions and opportunities in the working life of individuals. But it also contains rich data on other aspects of the social structures of European societies, for instance on the household and family patterns, the participation of individuals in central societal institutions such as education and labour markets, the distribution of economic, social and cultural resources among individuals (education, jobs of various characteristics, partly also income and wealth) and other crucial dimensions of living conditions.

Compared to other data bases the labour force surveys include several assets which enhance the strengths of the LFS for comparative studies:

• In most countries they include rather large samples of respondents which allow stable estimates even for rather selected social groups such as particular birth cohorts, workers with particular working conditions, the unemployed or special education groups.

• The surveys are replicated annually. The longer the series of these annual surveys becomes, the more valuable they are for the study of short or even long term social change.

• The data include detailed measurements of the various domains studied, in particular relating to jobs and the working life. It provides information in detailed classification schemas such as ISCO, NACE or the new ISCED classification.

• In most of the countries the design of the surveys include a panel component: In general the same individuals are observed at different points in time. A few areas of the questionnaire also include retrospective questions (e.g. on activities and jobs in the previous year). The new education module asks questions about the timing of events, such as the time of obtaining the highest educational qualification or the time of entering the
first stable job. All these elements allow the study of dynamic processes, at least for a few issues and in elementary ways.

• At least in principle, the design of the survey and the questions asked follow an agreed upon target structure that should produce sufficiently similar data in the various EU-member states and allow assessments at a high level of comparability.

The EULFS thus appears to be a really strong data base for the comparative study of many aspects of European Societies and it has a very high potential to contribute to a better knowledge and understanding of the similarities and differences of the societies and their converging or diverging change over time. Is it also a good data base for the comparative study of transitions from school to work - the topic of this conference? We will try to answer this questions in three parts:

• First, we briefly describe the core characteristics that need to be studied in the transition from school to work and how we understand the task of comparative research.

• Secondly, we give an example of a comparative analysis of one of the core issues in the transition from school to work - unemployment - and we evaluate the strengths and weaknesses of the EULFS to study this and other important issues in the area of school to work-transition research.

• Third, we comment briefly on promises and current limitations of using LFS data for transition research, which leads us into the concluding section, where

• Finally, we mention several crucial problems that urgently need to be solved for improving the quality of the EULFS and for extracting its full possibilities for the study of the transition from school to work in particular and its potential for the other areas mentioned above in general.

1. Transition from school to work and its comparative study with EULFS data.

Let us start with a few preliminary remarks on the issues involved in the comparative study of transitions from school to work.

As it has been observed by others, notably also by Michèle Mansuy and Patrice de Brouker, in present times for many young people the transition from school to work is not a single step that connects leaving the educational system and entering the world of work. Very often it is rather a process extending for quite a while and a process that often has neither an obvious starting point nor a clearly defined end. It may involve several steps forth and back between education and work. It may be interrupted by search and waiting times, involuntary unemployment or chosen time off for leisure, travel or other preoccupations. Also different states may be combined at the same time such as education and work and attachments to education or work may be more or less stable and intensive. All these activities may be taken for different reasons and therefore a similar objective state of, for instance, non-activity may have different subjective meanings for different people. Also a similar activity or non-activity
in different countries may have different subjective or objective implications because the context in which it takes place can vary. Therefore, it would seem that the generally cross-sectional character of labour force survey data and their enumeration of a few objective characteristics of states may not be particularly adequate to study the process of transitions in youth.

In fact, some crucial issues of interest can not be addressed in adequate ways due to the lack of information, in particular about the sequences of events and their subjective perception and evaluation. We will discuss a few of these issues further below. However, as our examples will show, there are nevertheless many topics for which even the stock data can provide useful information. To some extent the lack of flow information can be compensated by using the technique of pseudo cohorts and by imposing simplifying assumptions about the timing of events.

The comparative analysis of the LFS-data comes in mainly two variants. One is primarily descriptive. It is the social indicator approach adopted in a lot of the OECD work, for instance in the „education at a glance“-volumes or in the INES-project presented by Patrice de Broucker. If such indicators are defined in a theoretically meaningful way and if their measurements have an empirically unequivocal interpretation they can provide a powerful instrument in depicting important commonalities and differences between the countries. The major strength of the indicator approach is that in parsimonious ways countries can be described in a variety of dimensions and perhaps categorized according to a limited set of encompassing types. Many OECD- and EUROSTAT publications have shown that the labour force surveys can be used very forcefully to produce important insights about the realities of youth transitions in Europe.

There are, however two limitations to this approach: One is, that typologies and the allocation of countries to types may be rather sensitive to the chosen indicators. Furthermore, the observed measures for the indicators often are the outcome of the operation of various factors. In such cases, the substantive or theoretical interpretation of the measures found will be difficult. Descriptively, it is certainly interesting to know, that the level of unemployment in the transition stage is two or three times as large in the Southern European countries than for instance in Denmark, but within the indicator approach it is not easy to determine whether this is due to a different level of unemployment at large, or due to a different composition of young people in terms of qualifications with varying risks of unemployment, or due to some country specific institutional conditions that lead to a concentration of unemployment among labour market entrants or due to still other factors.

The second strategy of using LFS-data is more analytical. It attempts to disentangle and possibly quantify the factors, processes and mechanisms that lie behind the outcomes measured by the indicators. In this approach we need theoretical models for the various levels of social reality that specify which are the crucial variables and how they are related to each other. If we are successful we come to a deeper understanding of the phenomena observed
and we are able to indicated why the countries differ from each other in some observable outcomes. We will know whether the basic processes are similar in all countries or whether they vary depending on some specific contextual conditions. In this approach one will attempt to measure explicitly all conditions considered relevant and - following the recommendations of Przeworski and Teune (1970) - one will try to replace the names of countries by measured variables. Do the labour force surveys provide adequate data that can be used profitably in this kind of comparative research strategy? One evident issue is whether they provide measures for the variables needed to construct and empirically estimate theoretical models that are not inexcusably naive simplifications of reality.

As always: The proof of the pudding is in the eating. Therefore, in the next section we briefly present an example of such a model and analysis.

2. Unemployment Risks in the Early Career Stages: A Research Example
Rising unemployment risks for those entering the labour market in Europe are among the prime concerns of social scientists as well as policy makers in the fields of both education and labour markets, notably in the leeway of the recession in the early 1990s. On the other hand, there has been considerable cross-national variation in youth unemployment rates, persisting over and above any recent upwards trend: some European countries like Austria, the Netherlands or Germany are regularly found to perform much better in this respect than France, Spain or Italy, for example (European Commission 1997; Eurostat, various years; OECD 1996, 1998a). In order to improve our understanding of the problem, the Labour Force Survey offers a rich and readily available cross-national data source to describe these developments and to probe more in-depth explanations, which should ultimately allow for informed policy initiatives. The following examples intend to demonstrate these uses based on results from the ongoing CATEWE research project.

Describing Variation using the LFS: A Question of Levels or Processes?
Variation in unemployment rates abounds. As is very well known, there is huge cross-national variation in terms of unemployment risks at entering the labour market between European countries (OECD 1996, 1998a). Table 1 below provides evidence on this variation between 12 current EU member states, based on a specific definition of market entry. The concept applied introduces a career perspective on the transition process as market entry is defined relative to the date of completing one’s highest qualification, rather than biological age. So far, this date of completing one’s education has to be proxied by OECD’s (1997) estimated typical graduation ages for given levels and types of education in different countries; fortunately, the introduction of the year 2000 topical module into the EULFS will markedly improve the accuracy of information at least for this year. For convenience, those individuals within their first five years in the labour force will be defined as market entrants in most of the following. The results given in Table 1 reproduce many of the already well-known facts: young people experience much smoother transitions in countries like Austria (unemployment rate of 7%), the Netherlands, Denmark, and Germany (each rate around 10%) as compared to e.g. France
(28%), Greece (30%), Spain (37%) or Italy (42%). On the other hand, unemployment risks are very unequally distributed among skill groups. In addition to the cross-national variation in levels of unemployment, there is equally wide variation within each member state between leavers from different educational backgrounds. Typically, and particular exceptions notwithstanding, unemployment risks fall with increasing levels of education, as well as with occupational specialization in school-based vocational training, but more so by completing apprenticeships or similar forms of training.

Table 1  
Unemployment Rates among Market Entrants by Country and Education

<table>
<thead>
<tr>
<th>Country</th>
<th>Overall</th>
<th>Lower secondary</th>
<th>Dual system training</th>
<th>Upper secondary vocational</th>
<th>Upper secondary general</th>
<th>Lower tertiary</th>
<th>University degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>.068</td>
<td>.141</td>
<td>.057</td>
<td>.043</td>
<td>.103</td>
<td>.059</td>
<td>.049</td>
</tr>
<tr>
<td>Belgium</td>
<td>.157</td>
<td>.376</td>
<td>.140</td>
<td>.183</td>
<td>.232</td>
<td>[-----</td>
<td>.066 -</td>
</tr>
<tr>
<td>Germany</td>
<td>.104</td>
<td>.300</td>
<td>.093</td>
<td>.167</td>
<td>.090</td>
<td>.053</td>
<td>.053</td>
</tr>
<tr>
<td>Denmark</td>
<td>.105</td>
<td>.198</td>
<td>.100</td>
<td>[-----</td>
<td>.108 -</td>
<td>.057</td>
<td>.061</td>
</tr>
<tr>
<td>Spain</td>
<td>.365</td>
<td>.435</td>
<td>N/A</td>
<td>.338</td>
<td>[-----</td>
<td>.335 -</td>
<td>.265</td>
</tr>
<tr>
<td>France</td>
<td>.279</td>
<td>.490</td>
<td>.208</td>
<td>.254</td>
<td>.354</td>
<td>.173</td>
<td>.177</td>
</tr>
<tr>
<td>Greece</td>
<td>.299</td>
<td>.256</td>
<td>N/A</td>
<td>.303</td>
<td>.365</td>
<td>.285</td>
<td>.226</td>
</tr>
<tr>
<td>Ireland</td>
<td>.212</td>
<td>.408</td>
<td>N/A</td>
<td>N/A</td>
<td>.194</td>
<td>.104</td>
<td>.055</td>
</tr>
<tr>
<td>Italy</td>
<td>.423</td>
<td>.424</td>
<td>N/A</td>
<td>.351</td>
<td>.480</td>
<td>.518</td>
<td>.361</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>.097</td>
<td>.187</td>
<td>.099</td>
<td>.079</td>
<td>.143</td>
<td>.060</td>
<td>.092</td>
</tr>
<tr>
<td>Portugal</td>
<td>.129</td>
<td>.132</td>
<td>N/A</td>
<td>.161</td>
<td>.177</td>
<td>.118</td>
<td>.073</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>.172</td>
<td>.268</td>
<td>.158</td>
<td>.173</td>
<td>.136</td>
<td>.080</td>
<td>.082</td>
</tr>
</tbody>
</table>

Notes: Market entrants are defined as individuals in the first five years of labour market careers. Sources: European Union Labour Force Survey, pooled 1992-1997 database.

All of this is very well known and repeated in virtually each annual report by major statistical organizations and international organisations. Simple tables like the above do, however, allow for only limited inferences about the underlying mechanisms generating the level and distribution of unemployment risks in these different countries; hence, reliable guidance for policy concerns is equally limited. Of course, the ideal remedy to the any labour market research is regularly conducted longitudinal surveys of individuals combined with repeated cross-sectional surveys of (their) employers’ recruitment practices and demand for labour. Not even one part of this ideal labour market database is near available in comparative format. Given the absence of this “ideal” database for transition research, the CATEWE group has attempted to push the analysis of LFS data as far as possible. Two pieces out of this research will be presented extremely briefly here.

Of course, alternative approaches to statistical data analysis come at the price of introducing additional assumptions in order to identify more complex models or descriptive concepts. There is no statistical statement free of any assumption, however, so the issue is the
plausibility and usefulness of assumptions rather than their mere existence. Any cross-
sectional analysis ultimately has to assume the presence of equilibrium, and the descriptive
results presented in Table 1 moreover had to assume that OECD's proxy measures for typical
graduation ages indeed provide a sensible approximation to (the socially structured part of)
this individual level process. If one is moreover willing to assume negligible cohort effects
between adjacent annual cohorts of young people having entered European labour markets in
the 1990s, then LFS data might begin to tell more interesting stories about transition
processes: in that case, the available series of snapshot information from different annual
cohorts can be combined to yield a picture of unemployment risks over the initial years in the
labour market (based on the so-defined „pseudo-cohorts“). The outcome of such an exercise is
depicted in Figure 1, which gives unemployment rates among leavers from upper secondary
education (ISCED level 3) over the first ten years in the labour market after leaving the
education and training system. Leavers from upper secondary education should provide a
most interesting test case for cross-national variation as European educational systems
institutionally vary substantially at that level: in some countries, most leavers would come
from apprenticeships, while in other countries most leavers would have had school-based
vocational training, while yet others show large proportions of young people from
academically oriented tracks.

The figure holds many intriguing descriptive findings. First of all, there a huge level
differences right at the outset of careers for this skill group between European countries.
While unemployment rates are at a low 10% in Austria, Denmark, Germany, and the
Netherlands, rates in other countries amount to 20% in the United Kingdom or Portugal, to
around 30% in France, Belgium and Ireland, and even to more than 40% in most Southern
European countries. But the really interesting result is the fast convergence towards relatively
similar levels of unemployment for this skill group over the first years in the labour force.
After ten years on the labour market, cross-national differences are actually very much
reduced. In other words, the institutional environment of countries like Austria, Denmark,
Germany, and the Netherlands appears to be most suitable to achieve the inherent
„equilibrium“ unemployment rate for this skill group already at the outset of careers, while
convergence towards this „equilibrium rate“ takes substantially more time in other
institutional environments. That is, differences in levels of unemployment in initial working
lives might to a large extent reflect the operation of different institutional processes of labour
market integration – and lacking more adequate longitudinal data, the LFS is one of the few
comparative data sources which might be used to provide evidence on such differences in
integration processes. This is not to say that assumptions made in doing so are rendered
unquestionable thereby; rather, under the restrictions of these assumptions, LFS data hold at
least the potential to identify and sensitize the research and policy community for sources of
cross-national differences in transition processes, even as longitudinal data are most often
lacking.
Figure 1  Unemployment Rates by Potential Labour Force Experience (ISCED 3 leavers)

Notes: Lines represent smoothing of original estimates by logarithmic functions.
Explaining Unemployment Levels: A Multivariate Micro-Macro Modelling

The second potential use where LFS data has been by and large underrated is the area of comparative multivariate analysis. This is surprising indeed as the impact of a multitude of individual characteristics (like gender, ethnicity, education, family background) and local or national context factors (like local labour market conditions, aggregate economic situation, demographic changes, institutional regulations in training systems and labour markets) on transition outcomes is readily acknowledged in a growing body of research. Again, the EULFS as a standardized, large-scale and regularly repeated survey instrument which is run in a quite extensive set of countries appears like a natural choice for such analyses: the survey offers a readily available source of information on individuals’ labour market activities in very diverse regional and national contexts, in many cases even offering short time-series of aggregate observations. In conjunction with recent advantages in panel data econometrics and multilevel estimation techniques should increasingly enable researchers to exploit both micro- and macrolevel variation in their substantive accounts of transition (and other labour market) outcomes.

Along these lines, OECD (1998a) conducted a first multivariate analysis of unemployment risks in the transition from education to work, which was based on micro- and macrolevel controls gained from LFS data. In similar spirits, yet systematically using multi-level modelling techniques, three CATEWE papers adress a set of questions for comparative analyses of transitions into working life (van der Velden and Wolbers 2000; Gangl 2000a, 2000b). In one of these papers (Gangl 2000a), market entrants’ unemployment risks were related to the role of educational qualifications on the micro level, and institutional and socio-economic context factors at the (national) macro level. The analysis was primarily interested in the extent to which the effectiveness of similar qualifications in securing smooth transitions varied across institutional contexts, adequately controlling for measured and unmeasured context conditions at the country level as well as for unmeasured peculiarities surrounding specific national qualifications. Table 2 provides more extensive details on the structure of the model applied there.
In brief, the main outcome of the analysis was a close similarity in the effects of education among Northern Europe with some slight advantages to at least upper secondary level qualified individuals in those countries operating strong occupational labour markets. In contrast, the institutional environment of Southern Europe was found to work markedly against the immediate integration of better qualified leavers, while the least qualified experienced similar unemployment risks than their Northern European counterparts. These multivariate results finally also allowed to decompose cross-national differences in market entrants’ unemployment risks into several institutional and other components: (a) differences in the qualification „output“ of educational systems, which relate differences in transition patterns to institutional differences in education and training systems; (b) institutional differences in the evaluation of market entrants’ qualifications on the labour market; (c) country-specific elements of labour markets and the education-attainment linkage, and (d) systematic effects of the national macroeconomic context, among others. The results of this decomposition are given in Table 3 below.
Table 3
Institutions or Macroeconomic Context as Explanations for Cross-National Differences?

<table>
<thead>
<tr>
<th>Unemployment Risks among Market Entrants</th>
<th>Average Prediction</th>
<th>System Differences</th>
<th>Empirical Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(a+b)</td>
</tr>
<tr>
<td></td>
<td>Educational Systems - Educational Distribution</td>
<td>Institutional Labour Market Context</td>
<td>Educational Systems + Labour Market Context</td>
</tr>
<tr>
<td>Austria</td>
<td>0.202</td>
<td>-0.035</td>
<td>-0.046</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.202</td>
<td>-0.026</td>
<td>-0.014</td>
</tr>
<tr>
<td>Germany</td>
<td>0.202</td>
<td>-0.047</td>
<td>-0.046</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.202</td>
<td>-0.026</td>
<td>-0.046</td>
</tr>
<tr>
<td>Spain</td>
<td>0.202</td>
<td>+0.025</td>
<td>-0.014</td>
</tr>
<tr>
<td>France</td>
<td>0.202</td>
<td>-0.009</td>
<td>-0.014</td>
</tr>
<tr>
<td>Greece</td>
<td>0.202</td>
<td>+0.008</td>
<td>+0.115</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.202</td>
<td>+0.013</td>
<td>-0.014</td>
</tr>
<tr>
<td>Italy</td>
<td>0.202</td>
<td>+0.041</td>
<td>+0.115</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0.202</td>
<td>-0.013</td>
<td>-0.046</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.202</td>
<td>+0.049</td>
<td>+0.115</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.202</td>
<td>+0.012</td>
<td>-0.014</td>
</tr>
</tbody>
</table>

Notes: Predicted system differences compared to the average prediction; all predictions are based on a multivariate random effects logit model based on the pooled 1992-1997 European Union Labour Force Survey controlling for gender, potential labour force experience, level and type of education as microlevel determinants of labour market success but also macrolevel factors like institutional labour market context (occupational labour market systems, Southern Europe, other European countries) and various macroeconomic conditions (aggregate economic situation, youth cohort sizes, level of educational expansion and professionalization tendencies in the labour force), allowing for unmeasured heterogeneity between countries and types of education in the estimation.

From this analysis, the crucial importance of institutional factors for the explanation of cross-national differences in unemployment risks at entering the labour force clearly emerges. From the column (a+b), which reports the simulation results for institutional differences combined with the compositional differences in terms of market entrants’ educational backgrounds, the institutional contribution to the positive performance of countries like Austria or the Netherlands and the less favourable outcomes in most of Southern Europe may be gauged. According to these figures, systematic institutional differences are able to explain between 60%-70% of the empirically favourable performance of the famous set of countries that operate strongly vocationally oriented training systems. These institutional effects are related to both the existence of fairly large-scale systems of apprenticeship training (which has favourable microlevel effects) and the more favourable transition patterns among better qualified leavers in these systems (i.e. there - relatively - enhanced competitiveness as compared to more adult workers). In Southern Europe, in turn, very high unemployment risks clearly emerge as an institutional problem in the labour market. In sum, while there are of course macroeconomic context effects on youth’s employment chances, there is a large role for institutions to play in bringing about the observed cross-national differences. Obviously,
these results do not yet allow to precisely pin down the institutional sources of differences in labour market performance in all cases; but they seem to at least suggest a viable approach for further study.

3. Transition Studies using LFS: Survey Potentials and Limitations

Given the wealth of basic information on labour market activity contained in the LFS, LFS-based transition studies are, of course, not restricted to the domain of unemployment risks. Within the CATEWE project, different analyses have e.g. explored cross-national variation in the educational backgrounds of market entrants (Müller and Wolbers 1999), cross-national variation in the incidence of double status situations, i.e. the co-occurrence of participation in training activities and working, in the transition period (Wolbers 2000), cross-national variation in the incidence of non-standard forms of employment like part-time or fixed-term contracts (Couppié and Mansuy 1999; Wolbers 2000), cross-national differences in young people’s industrial allocation patterns (Couppié and Mansuy 1999), or cross-national differences in terms of initial occupational allocation (Gangl 2000a, 2000b) and their evolution over the early labour market career (Gangl 2000c). These studies did not even begin to touch upon important matters like cross-national variation in terms of gender or ethnic differentiation in youth transitions, the impact of regional and local economic conditions on early labour market outcomes or the interaction between labour force, household context and family (formation) behaviour, all of which could potentially be addressed from existing LFS material. Even more so, the introduction of the ad-hoc topical module on transitions in the LFS 2000 will allow to study the point of leaving education and training more precisely, to explore occupational mobility over the early career stage at the individual level and to address the impact of social origin on education and transition outcomes of young people. Moreover, the potential of the database for feeding into studies of both short-term processes using pseudo-cohort-techniques (OECD 1996; Gangl 2000c) and longer-term studies of cross-national variation in patterns of social change by true cohort analyses (Müller and Wolbers 1999) are only about to unfold.

To be sure, the (EU)LFS is far from being a perfect source on transitions processes from school to work: there are (and probably always will be) problems with how certain concepts are applied, which might be fitting for one purpose but less so for another; with restrictions in terms of variables included, which might be appropriate for some questions but fail to cover important aspects of certain more specific issues; or with the basically cross-sectional orientation of the survey which yields only limited information of labour market flows. More concrete examples of LFS limitations – gained from our research experiences in using LFS for transition research – would include:

- severe access restrictions to the EULFS data, preventing effective use of the full wealth of information contained in the database;
• difficulties in defining the timing of leaving the education and training system and, equivalently, the timing of (first) entry into the labour market; multiple labour market entry is not identifiable;
• validation of data harmonization/standardization processes neither possible from the data nor documented (e.g. education variables);
• impossibility of using household context information, notably partner labour market activities, as household records can typically not be identified;
• impossibility of using the LFS panel component (even as a panel of locations as administered in some countries) as individual survey units cannot be identified;
• availability of only categorical duration measures (unemployment duration; tenure)
• restricted availability of retrospective information; e.g., occupation, earnings and type of contract unavailable for employment positions held one year before;
• practically no availability of more subjective information on e.g. labour market attachment; for specific subsamples, i.e. women re-entering the labour force, more extended retrospective components covering non-market activities might also be helpful

Certainly, it is important to recognize that adaptational changes in a large-scale survey instrument like the LFS are limited almost by definition as they cannot and should not overburden the instrument or fundamentally re-orient the survey according to short-run concerns. The strength of the LFS as a basic source on European labour markets lies in the reliability of the instrument gained from long years of experience with it and the reliable repetition of a core set of questions over an extended period of time. These advantages should not easily be wasted. Therefore, it seems likely that further piecemeal adaptation like the intention to use quarterly designs in all European countries or the introduction of (ad-hoc or more regular) topical modules imply promising further developments of a well-established instrument. On the one hand, quarterly rotating periods will provide greatly enhanced details on labour force dynamics, while topical modules can serve to flexibly adapt the instrument to short-run interests. For example, the issue of defining individual dates of leaving the educational system will be solved by the Eurostat ad hoc topical module for this year (with the insights gained from this additionally enabling for better proxy measures in other survey years). On the other hand, it is probably unnecessary to administer very specific modules at annual intervals as major social change will not have occurred in between.

What has to be recognized more explicitly, though, is that current problems with using LFS as a source for transition and other research, come at various levels, with the conceptual level addressed above being only the most distant one. Indeed, the potential of LFS analysis could be enhanced markedly if only full use of the already existing information could be made. A case in point might be the analysis of the joint labour market behaviour of couples. As long as households or families cannot be identified from the individual records, it is per sé impossible to address such joint decisions in households. It is obvious that any separate and individual account of male and female labour market activities is theoretically implausible and substantively unconvincing. As LFS is an household survey, the only information necessary
to provide to the scientific community for such analyses is an anonymised household level ID. Other similar examples, which relate e.g. also to survey documentation, may easily be taken from the list provided above or found in other studies. We therefore close this paper with some considerations for further improvement of the survey from the point of view of empirical social research.

4. Suggestions for improvements of the data base and its availability for Scientific Research

All in all, the examples discussed aptly demonstrate the potential of the European Community Labour Force Survey (ECLFS) as a database for comparative transition research. Actually, this study is the only database which provides comparative microdata on labour force behaviour and employment covering an extensive time series for up to 15 European countries. Given the strong potential of the database and given the fact, that we have extensively used it for our work, it may also be useful to discuss specific problems of using the EULFS database in comparative transition research. Eurostat, as the collector and provider of the database, is well known for both the quality of data processing and statistical knowledge, as well as its strong orientation towards the constant improvement of one of the main European data sources. Specifically in the area of transition research, an extensive responsiveness to concerns expressed in scientific research is obvious: from the experiences of the current analyses, efforts made by Eurostat to introduce a regular topical module on transitions from education to work into future Labour Force Survey studies or to supply an individual status definition accompanying the traditional ILO concepts of employment, unemployment and labour force participation are much appreciated and likely to enhance the potential of the database substantially. Acknowledging thus both the high quality standards achieved and the extensive content already covered in the EULFS database, we would like to add some suggestions for even further enhancement of the database quality and their use to scientific research on European issues.

Provide more extensive documentation on the EULFS database

From a purely pragmatical point of view, the fully appropriate use of the EULFS database is often inhibited due to sheer lack of documentation. Undoubtedly, Eurostat’s „Methods and Definitions„, series is invaluable in presenting the outlay and contents of the database. However, applying the EULFS database in comparative research regularly raises issues which cannot be answered from this series, yet are often crucial for an adequate consideration of the information provided by the database. Examples comprise:

(a) Documentation of national procedures for variable generation
The power of the EULFS database lies in the provision of standardised, comparative information on labour force behaviour and employment for currently 15 European countries. Still, the power of the database is crucially affected by the validity precisely of the standardisation processes adopted. Standardisation is likely to be less of an issue as long as standard ILO concepts of employment, unemployment, labour force participation and the like
are concerned. The survey does, however, contain a number of variables which are covered by international definitions, yet where the processes and set-ups of national proceedings to meet such a definition are not immediately obvious. Nevertheless, the value of comparative analyses precisely lies in establishing cross-country similarities and differences from comparable data – so comparability and validity of data definition are crucial to the endeavour.

As European educational systems vary widely, information on individuals educational background is an area where such problems are likely to arise. Consequently, the „Methods and Definitions„ series provides a list of national mappings of qualifications and credentials to the common ISCED concept. Still, the listing is hardly more than a translation for some countries, it is only available since the introduction of the 1992 LFS series, and, finally, not all changes introduced in the codings are documented even from then on. Other examples of extremely interesting information provided by the survey but without sufficient information concerning the comparability and the validity of national codings include: the definitions for respondents’ status one year ago (specifically, how apprentices are treated), the definitions of the type of education and training currently received (which training is dual system training, which training is in-company training only?), or the purpose of current studies. In all of these and related cases, clear and complete national information on what is actually represented in the EULFS codes is required for appropriate comparative analyses in much the same way as has been introduced in the „Methods and Definitions„ series in the case of information on respondents’ educational background.

Similarly, it is indispensable to know about any modification or change introduced in the codings of variables, for example after a scientific evaluations of national mappings used or other national changes in coding schemes, even if unrelated to a general major revision of the LFS itself.

(b) Documentation of survey sample details
Apart from such definitional issues, the responsible analyst constantly lacks very simple information which, however, crucially influence the assessment of the reliability and quality of the statistical estimates gained in analysis. In any statistical analysis, it is e.g. indispensable to know the unweighted size of the survey sample in order to be able to calculate appropriate hypothesis tests. Equally well, it is of crucial importance to know precisely about the survey target population and the calculation of population weights supplied.

Of course, not all of such information will be needed in every kind of analysis, and, consequently, not every researcher will be interested in all such details. For most information purposes, researchers are well served with the „Methods and Definitions“ series. Still, for many instances the documentation of the full study details would be very useful and are very much needed. Therefore, it is proposed to introduce a „Technical Reports” series which would cover such information as the calculation of weights or the precise national
codings for specific variables in the European Community Labour Force Survey, which can be referred to by interested researchers if such information is needed.

Provide more extensive access to database for scientific research

The power and potential of the EULFS database in providing comparative information to analyses of European interest can only unfold fully if the scientific community can, under sufficient regulation, gain full access to the database. The more restricted access to the database, the less it is possible to draw an adequate picture of European labour markets and, consequently, to direct political action where necessary. This is certainly not to deny the importance of data protection legislation: Confidentiality of individual information is of indispensable and unquestionable primary interest. Still, the present opportunities and conditions of using the database for scientific research is far from being satisfactory and would be much improved without any serious threat to data protection.

Currently, access to the EULFS is restricted mainly by the number of variables and cell sizes in the multi-dimensional table requested: if weighted cell sizes regularly fall below 50 persons access to the requested information is denied. This may be a very sensible restriction in statistical terms against unreliable population estimates. Still, the distinction between statistical and data protection necessities is crucial. The EULFS data itself is already subject to national data anonymisation procedures; as such, the potential of identifying individuals from the EULFS is no larger than doing so from national source data. In many countries the respective national individual level microdata are available to the research community. The scientific interest is in modelling the distribution of employment, unemployment, qualifications, wages and the like, depending on sets of explanatory variables. The most flexible and fruitful approach to this simply is to be able to process anonymised individual level data, allowing both for detailed validation of data definitions and appropriate statistical modelling. For using the full potential of the ELFS for scientific research, it is urgent to find appropriate ways to make anonymised individual level data accessible to the research community as soon as possible.

To judge from our own experience: Most if not all analyses that we have done in the CATEWE-project, could have been more extensive, more detailed and even more convincing if it had been possible to include one, two or three more variables into the study. Therefore, ways and means should be found to be able to provide access to the full EULFS database to interested scientific researchers, both in terms of individual records and the full set of variables.

Such a solution may involve time consuming negotiations with the EU-member states who collect and deliver the original data to EUROSTAT. It may also take time to devise appropriate procedures for anonymising the data. While waiting for an adequate solution in terms of a scientific use file, there might be other ways to give interested and competent researchers access to the full power of the database: Researchers might be enabled to process their own programs during a stay at Eurostat. Alternatively, considering the example of the
Luxembourg Income Study collection, it could be possible to enable registered researchers or research projects to submit statistical programs to a central server at Eurostat, taking care for the processing of these jobs drawing on individual records, yet delivering only the statistical output back to researchers. What is important for such a solution is that researchers can obtain a small file of randomly selected and anonymised cases from the full data base in order to be able to test their programs.

No matter how the issues raised here can be solved pragmatically, the claim is that much of most interesting and relevant scientific research into labour market processes needs more of the individual level data than current practices do yield.

*Provide access to the harmonised data base as well as to the original national data files*

For many purposes the inputs of Eurostat and of the national statistical offices to construct a harmonised database are extremely valuable. Studies like ours that could draw on harmonised data covering a period of several years for a large number of countries would never have been possible without this input. However, any construction of a harmonised data base always involves decisions that may be adequate in one theoretical perspective, but may be less satisfactory under a different perspective or in view of different research questions.

This is particularly evident with respect to education. Given the institutional complexity of European educational systems there is not one single way how to classify in a comparable way for instance educational attainments in the various countries. Researchers should therefore be enabled to construct alternative harmonised classifications if they have reasons to do this (or if they are hazardous enough to engage in such an endeavour). Let us give a brief example that it may be worth while and important to draw on the original national classifications and that the constraints involved by constructing harmonised data may lead to misleading conclusions.

In harmonising educational information the EUROSTAT, OECD and other international organisations generally use the ISCED-classification. In the harmonised EULFS database educational information is also coded according to the ISCED-rules. On of the results regularly found from data coded according to these rules is, that in most European countries women meanwhile outperform men in terms of educational participation and attainment, in particular also on the tertiary level. In the Education Key Figures- publication (European Commission, 1997), Eurostat for instance, already in 1997 states:

„Twenty years ago, women were underrepresented in tertiary education in all member states. Over time, the proportion of women has continuously been growing and in most member states they represent the majority of students nowadays.“ (translated from the German language document, WM/MG)

Indeed, as figure F7 in European Commission (1997) shows, more women than men participate in recent years in educational courses assigned to ISCED level 5-7. The OECD-
indicators on the proportions of the population that has attained at least an university-level education show a similarly positive picture about the virtual disappearance of gender inequalities in educational participation in many countries. For the age group 25-34 years OECD (1998b:45) notes that in 1966 e.g. in France 12% of men have obtained an university-level education, while among women the number is 13%. On contrary, for Germany and the UK the OECD finds still a slightly larger proportion of men than of women with university-level qualifications.

Figure 4. Proportion of women compared to proportion of men among graduates of different levels of tertiary qualifications by cohort

* Proportion of women among graduates from given level of qualification compared to proportion of all women among birth cohort considered

Source: Own calculations from microdata of national Labor Force Surveys of Germany (1993), France (1994) and United Kingdom (1994)
The original national classifications, however, for all three countries provide a more differentiated picture, and at the same time show a less favourable position for women in all three countries. If one differentiates, as in Figure 2, the heterogeneous category of tertiary education graduates along more specific educational degrees, then inequality between genders is still clearly evident. In England and France, women more often than men obtain those tertiary degrees which are less prestigious, less demanding and provide less advantageous opportunities on the labour market. Men, on the other side, are over-represented among the degrees at the top of the education hierarchy, even in most recent times. Germany is characterised by the fact that for both levels of tertiary qualifications the proportions of women are lower than those for men. Similarly differentiated analyses would probably show continued gender inequalities on the tertiary level also for other countries.

Develop Future LFS Surveys, Enhance the Analytical Potential of Earlier Surveys

The power of the EULFS database lies in the provision of standardised, comparative information on labour force behaviour and employment for currently 15 European countries – for some countries and variables in time series format for up to 15 years by now. That is, the EULFS database not only provides an excellent tool for comparative research on current European labour markets, but rather is a unique source for analysing and understanding labour market changes over the last decade in European perspectives. As the LFS survey design, survey methodology and survey content has naturally changed, evolved and improved over time, its use for comparisons across time is not without problems.

Already in the past, crucial variables like educational background, occupation or industry of current employment have been subject to substantial revisions within the lifetime of the EULFS database. And recalling the introduction of a topical module on transitions or the development of new scales in coming LFS series, Eurostat itself has proven to be interested in and capable of even further substantial improvements of the EULFS database. EULFS New elements are constantly introduced into the LFS survey: new questions are asked, new EU countries join the database, or recent revisions of international standard classifications are implemented. Undoubtedly, improving current and future surveys is of pre-eminent importance for most uses of EULFS data and with such improvements the potential use of the EULFS for scientific purposes is bound to grow markedly in the very near future. Nevertheless, understanding longer-term changes often is the key to understanding current cross-country similarities and differences.

In this line of reasoning, maintaining and, possibly even enhancing the time series potential in the EULFS database appears to deserve more serious concerns than have been apparent so far. Therefore, it is suggested here to complement efforts to improve current and future LFS surveys by directed and systematic action to investigate into, to keep up and improve the time series capabilities inherent in the EULFS.
Systematic work and research is needed to clarify the potential of time series use of the database by e.g. addressing potential reconciliation of coding changes introduced. This is a worthwhile task because no alternative database could provide similar insights. To our impression, reconciling the informational base incorporated in different LFS series yields reasonably stable time series information which enables to answer crucial questions about currently occurring changes in European economies. EULFS Definitional changes must not necessarily mean that information gained earlier is inevitably lost for scientific purposes. But in-depth investigation into the consequences of design changes implemented and the relations between current and earlier LFS designs must uncover and retain the time series potential of this unique database.

These proposals for further improvement of the survey do, of course, need further elaboration and discussion. Given the outstanding importance of the database for social scientific research in and on the European Union it is our conviction, that high priority should be given to investments for the further development of the EULFS database and for making it available to the research community in more adequate ways. Considering all the work that is involved in developing the EULFS to its full potential we were often surprised in the past how a small group of researchers at Eurostat must cope with all these issues and how often personnel in the EULFS task force is exchanged while in fact the tasks needs a profound professional expertise that can only be gained with continuous work and research in the area.

References


