

Integration of data from different sources: Unemployment

by I. Chernyshev*

1. Introduction

Recently, the ILO Bureau of Statistics began to study the use of unemployment data from different sources. The ultimate objective is to study the use of both employment and unemployment statistics collected from different sources and develop harmonized concepts and methods of data reconciliation.

Information on the social and economic situation in the labour market is mainly collected from the following three broad groups of sources:

- (i) population censuses and labour force surveys;
- (ii) establishment surveys; and
- (iii) administrative records/register-based statistics;

While these multiple sources of data provide users with a wide spectrum of statistics, they may sometimes create confusion as to which of them reflect more accurately and more completely a given phenomenon. Moreover, it is not uncommon that data collected for the same economic variable but from different sources may lead not only to different results but to show opposite trends.

Such discrepancies in labour market statistics stem from the fact that all the sources listed above have different primary objectives, different coverage, different reference periods and frequency; they also use different definitions and classifications of closely related concepts.

This article presents the first product of the ILO study: a table containing unemployment data collected from two types of sources - labour force surveys (LFSs) and administrative records (ARs) - in thirty-six countries. The results of a global comparison of data from the above two sources over the period 1975-1996 are presented below.

2. Tabulations

The data in Table 1 (Unemployment from different sources, 1975-1996) provide information on the number of unemployed - total and by sex - in thirty-six countries. The table was constructed on the basis of information available in the ILO *Yearbook of Labour Statistics, 1997* and prior editions, the ILO quarterly *Bulletin of Labour Statistics (1998-1)* as well as data provided by national statistical offices in reply to a special request. The latter initiative was taken in those cases where the time series published in the ILO publications were either incomplete or missing. As a rule, when a series was missing it meant that the data from that particular source ceased to be used as the official measure of unemployment in the country. However, the available data have been included in Table 1 for the purpose of this study.

* Bureau of Statistics, International Labour Office.

As only a few countries conducted regular labour force surveys prior to 1975 and their number started to grow dramatically afterwards, it was decided to begin the ILO exercise with that particular year and cover the next two decades. In fact, Table 1 shows that the second half of the 70's, the mid-80's and the beginning of 90's were the landmark periods of the massive introduction of LFSs in the national statistical programmes of many countries. Thus, while in 1975 there were only ten countries conducting regular LFSs, their number doubled in 1985 and more than tripled in 1995.

It is also worth noting that data presented in Table 1 are "raw" in the sense that they are not accompanied by notes or references and therefore should be used primarily for a global comparison rather than an in-depth comparative study and analysis.

3. Comparison in time

Main findings

A global comparison of the data presented in Table 1 reveals that through time the LFS total unemployment figures are always higher than those collected from ARs in twelve countries: Bulgaria, Canada, Czech Republic, Estonia, Greece, Latvia, Lithuania, Russian Federation, Sweden, Turkey, Ukraine and the United States. Conversely, in the following eight countries the total number of AR unemployed always remains higher than those of LFS unemployed: Belgium, Hungary, Ireland, Italy, New Zealand, Poland, Slovak Republic and Slovenia. The gap between the sources appears to have a varied pattern over time. Thus, the study shows that there is no country where such a gap remains stable over a sustainable period, in fact it increased in five, decreased in eight, showed reverse trends in sixteen and had an irregular pattern in seven countries. The above findings are valid for total unemployment only. As for unemployment by sex, the study demonstrates that once split, the figures show different patterns of behaviour from those levelled off by the totals. They may not only weave around or leap to and from a given ratio of totals but significantly deviate from it and sometimes even yield opposite results.

For example, while in New Zealand the LFS total unemployment has always remained lower than the total number of AR unemployed, there were five years when the LFS female unemployment was higher than the registered one. Conversely, in Sweden, the country where the LFS total unemployment has always been higher than the number of AR unemployed, in 1994-96 the registered female unemployment outnumbered the LFS female unemployed. In Finland, when in 1987 the total unemployment measured through the LFS and ARs was the same, the ratio of males was 5.4% and that of females -5.4%.

Different patterns

In the summary mentioned above, the longitudinal curves of the LFS/AR ratio of total unemployment reveal the following four major behavioural patterns over time: (a) increase; (b) decrease; (c) reverse; and (d) irregular.

(a) *Increase*: the gap between the two sources increased in the following countries: Canada, Ireland, Italy, Lithuania and Slovenia. This pattern is observed both in countries

with positive and negative ratios. Out of the four, the most spectacular increase was in Canada: from 23.9% in 1976 to 142.8% in 1996.

(b) *Decrease*: the gap between the two sources decreased in the following countries: Bulgaria, Czech Republic, Greece, Hungary, Latvia, Russian Federation, Sweden and Ukraine. This pattern is observed both in countries with positive and negative ratios as well. The most spectacular and consistent decrease was in Sweden: from 119.2% in 1977 to 1.0% in 1996, and in Greece and Hungary there was a bulge at the end of the period.

(c) *Reverse*: during the observed period, the LFS/AR ratio has changed from positive to negative and vice versa in the largest group of countries. They are: Austria, Denmark, Finland, France, Germany, Iceland, Israel, Luxembourg, the Netherlands, Norway, Portugal, Romania, Singapore, Spain, Switzerland, and the United Kingdom. These countries can further be broken down into the following three sub-groups: (i) countries where the positive ratio became negative: Denmark, Iceland, Luxembourg, Portugal, Switzerland;

(ii) countries where the negative ratio became positive: Israel, Romania, Singapore; and (iii) countries where the ratio changed through time poles repeatedly: Austria, Finland, France, Germany, the Netherlands, Norway, Spain and Switzerland.

The comparative analysis of group (c) shows that the most dramatic change occurred in Singapore, where between 1975 and 1996 the LFS/AR ratio of total unemployment rocketed from -5.0% to 4297.2%. It is followed by Romania where ratios abruptly swapped poles between 1994-1996 from -18.0% to 20.1%. The most gradual downhill slide is observed in Iceland, with ratios changing from 89.4% to -5.0% and a cross point at the end of the period, and Switzerland where they changed from 73.4% to -14.0 % with a cross point at the beginning of the period. The German ratios, after a decade of steady climb, plummeted in 1996 from 14.6% to -9.8%. In contrast, the Finnish ratio leaped during the same period from -4.7% to 10.6%, at the same time this is the only country where the total unemployment from the two sources was equal - this happened in 1985 and 1987. The Portuguese ratios, demonstrate an interesting case of almost symmetrical summits and downfalls registered between 1975 and 1996. Another interesting case is found with the Dutch ratios, which form a curve crowned with two humps resembling a roller coaster and reflecting the steep falls occurring between 1979-1985 and 1991-1994. While the curve of Denmark has numerous blips, in Austria, Luxembourg and Spain the gaps between the sources have the least expansions and contractions among the countries whose ratios dropped more than once below a zero point during the last decade.

The charts presented below show examples of different reverse patterns found during the study.

(d) *Irregular*: this pattern prevailed in the following seven countries: Belgium, Estonia, New Zealand, Poland, Slovak Republic, Turkey and the United States. Out of the seven, four countries have LFS/AR ratios with gaps fanning out and three with gaps fanning in. Belgium's curve has the sharpest edges followed by the United States and Turkey. New Zealand has the least disturbed curve while that of the Slovak Republic pitched in 1996.

4. Conclusions and future work

The study shows that in spite of their differences, the thirty-six countries covered have a number of similarities which make it possible to classify them into four major groups. This classification should be explored further on a country by country basis and, ultimately, used for the development of harmonized concepts and methods of data reconciliation.

It appears that the behavioural patterns of total unemployment differ from those formed by unemployment broken down by sex. Furthermore, it seems that any further breakdown of unemployment figures, say, by age, region, etc. will lead to new patterns of LFS/AR unemployment ratios. All this should be taken into consideration while developing a reconciliation frame.

The study also reveals that the size of the gaps between the sources has no direct correlation either with the length of time a country has been conducting a labour force survey nor with a country's tradition in keeping population registers. Thus, it has *not* been noted that the dispersion of figures in Denmark, Finland, Norway or Sweden are significantly smaller than in the Czech Republic, France, Germany or Hungary.

The data assembled in Table 1 should, however, not be used for a direct comparison of discrepancies between the two sources and even less so as a basis for criticism of the differences found. In fact, as has already been mentioned, some series ceased to be used at all by certain countries as the official measure of unemployment.

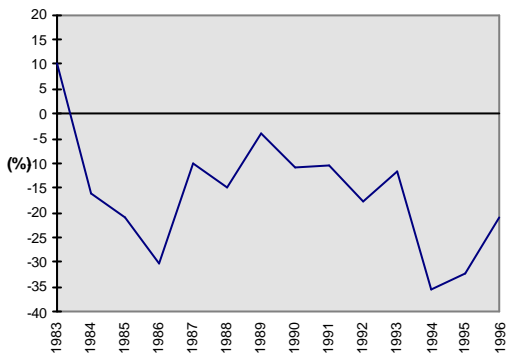
In order to understand why the data differ and what should be done to reduce the differences to an acceptable minimum, it is necessary to study carefully the reasons for discrepancies between the sources. The latter may be due to inconsistent definitions, different reference period, non-contiguous classifications, variable coverage, measurement errors, etc.

Future ILO work will be concentrated on: (a) expansion of the number of countries currently covered in Table 1; (b) country by country study of the national practices of the use of unemployment data collected from different sources; (c) in-depth country case studies of data discrepancies; (d) study of national experiences of data comparisons and reconciliations; (e) development of harmonized concepts and methods of data comparison and reconciliation/integration, in collaboration with national specialists; and (f) preparation of a glossary on international terminology for different steps of data comparison, reconciliation/integration.

It is envisaged that a similar study will be carried out on employment statistics from different sources.

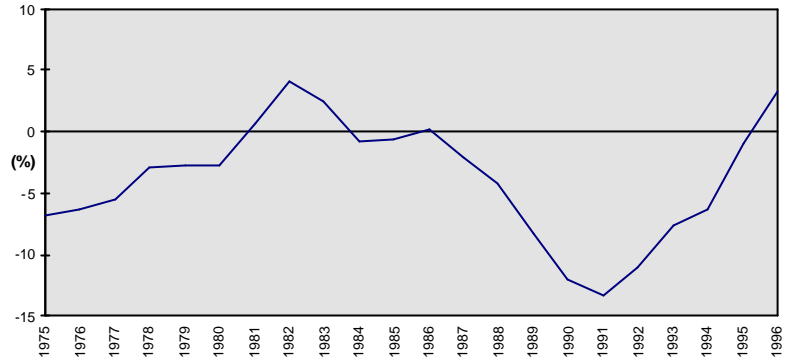
Examples of different reverse patterns found during the study: LFS/AR total unemployment ratio

LFS vs AR Unemployment: Total



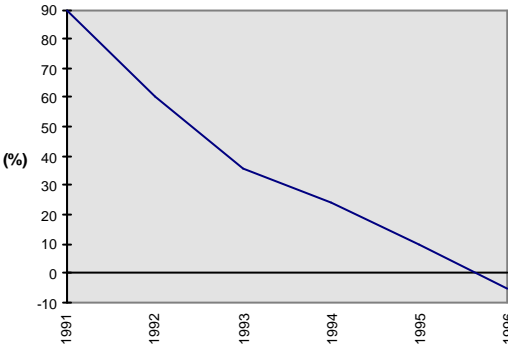
Denmark

LFS vs AR Unemployment: Total



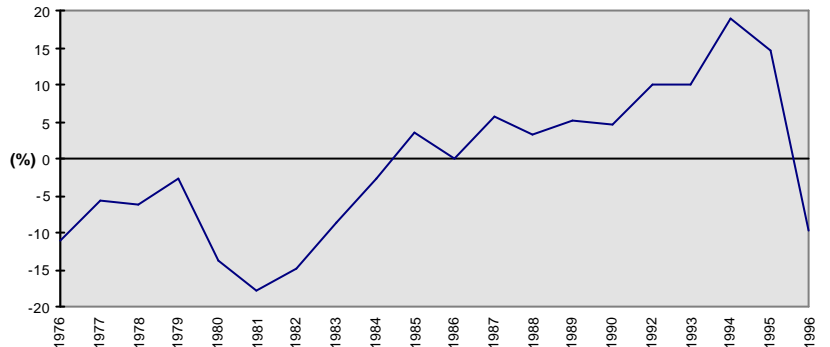
France

LFS vs AR Unemployment: Total



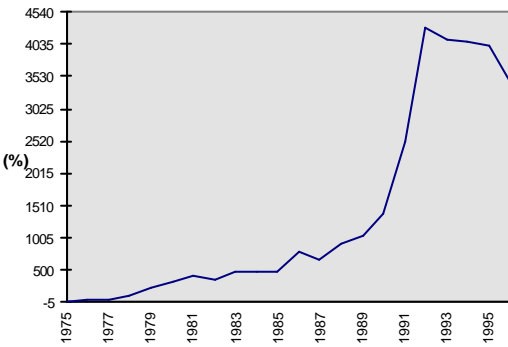
Iceland

LFS vs AR Unemployment: Total



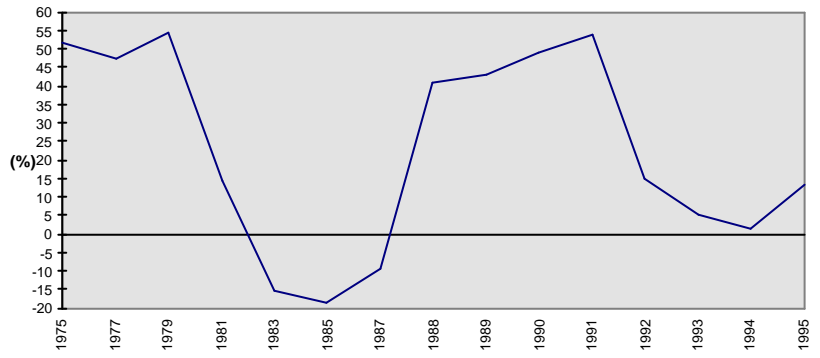
Germany

LFS vs AR Unemployment: Total



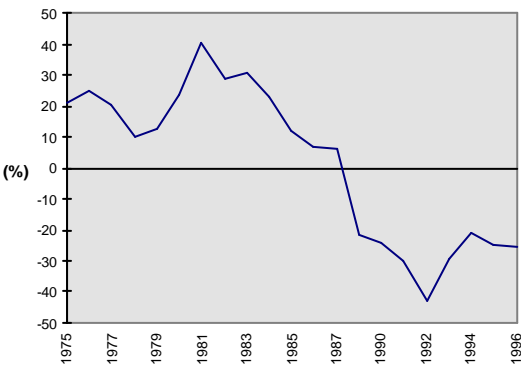
Singapore

LFS vs AR Unemployment: Total



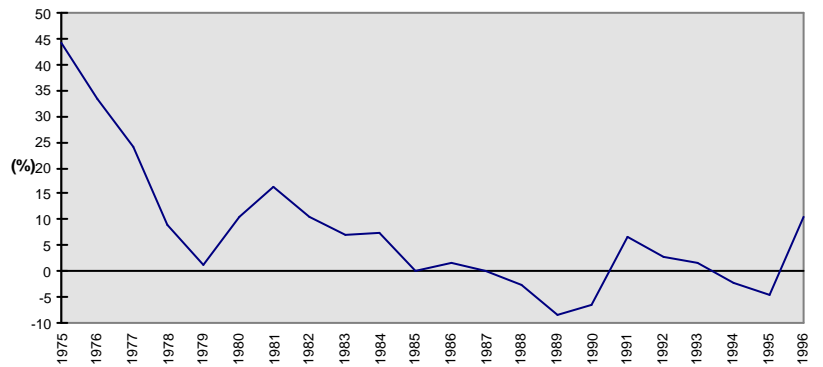
Netherlands

LFS vs AR Unemployment: Total



Portugal

LFS vs AR Unemployment: Total



Finland

