

Short term statistics quality Reporting: the LCI National Quality Report 2004

Fabio Massimo Rapiti()*

(*) ISTAT

Direzione Centrale delle statistiche economiche congiunturali su imprese, servizi e occupazione - Statistiche congiunturali sull'Occupazione e sui Redditi.

Riassunto

L'importanza delle diverse esperienze di documentazione sulla qualità delle statistiche.

Negli ultimi anni, a livello nazionale e internazionale, è divenuto sempre più importante preoccuparsi dell'accessibilità e della chiarezza della documentazione sulla qualità delle statistiche. D'altronde, la credibilità degli istituti nazionali di statistica dipende in modo cruciale dalla loro capacità di trasmettere agli utilizzatori e al pubblico fiducia sulla qualità delle statistiche prodotte. La possibilità per gli utilizzatori di avere accesso all'informazione sulla qualità delle statistiche diventa, allora, una sfida fondamentale per la statistica ufficiale.

Molti regolamenti dell'UE, ormai da alcuni anni, richiedono ai paesi membri di valutare e certificare la qualità delle statistiche attraverso specifici "Quality Report". Il fine di questi rapporti è di fornire all'Eurostat e alla Commissione Europea strumenti per valutare la qualità delle statistiche attraverso un insieme di indicatori qualitativi e quantitativi. Ciò è certamente molto importante ma non è sufficiente: per aumentare la trasparenza e rafforzare la fiducia degli utilizzatori nelle statistiche sarebbe necessario che le informazioni sulla qualità venissero messe regolarmente a disposizione del pubblico e fossero facilmente accessibili. Rapporti o documenti sulla qualità dovrebbero essere regolarmente pubblicati e aggiornati.

Verificare la qualità delle statistiche congiunturali è particolarmente difficile. Infatti la documentazione disponibile regolarmente sulla qualità degli indicatori mensili o trimestrali è piuttosto rara. Per esempio, nonostante la discussione su questo tema sia stata ampia e duratura nel gruppo di lavoro Eurostat sulle statistiche congiunturali sulle imprese, a tutt'oggi, per gli indicatori infrannuali non è obbligatorio fornire regolarmente rapporti sulla qualità. Soltanto alcuni studi di fattibilità e di carattere preliminare sono stati portati avanti.

Per queste ragioni sembra utile far circolare documentazione su quelle poche esperienze in cui la reportistica sulla qualità è stata introdotta. Il Regolamento sull'Indice trimestrale di costo del lavoro (n. 450 /2003) nell'art. 8 richiede che ogni stato membro fornisca all'Eurostat un Rapporto Annuale sulla qualità. Il rapporto presentato nelle pagine seguenti è stato realizzato a fine 2004 e rappresenta la prima esperienza condotta nella Direzione centrale delle statistiche economiche congiunturali sulle imprese, i servizi e l'occupazione e nel Servizio statistiche congiunturali su occupazione e redditi. Proprio per questo esso è ancora ad uno stadio molto preliminare ma, a nostro parere, sufficientemente completo. Anche per questa ragione è opportuno che circoli nella comunità scientifica interna ed esterna all'istituto nazionale di statistica. E' benvenuto ogni commento o feedback sulla struttura del Rapporto, sulle sue debolezze, sugli indicatori utilizzati. Per ogni informazione si prega di contattare Fabio Rapiti al seguente indirizzo di posta elettronica: oros-info@istat.it.

Il rapporto è stato trasmesso dall'Istat (Direzione centrale delle statistiche economiche congiunturali sulle imprese, i servizi e l'occupazione) all'Eurostat nel mese di dicembre 2004.

Summary

The importance of exchanging experience in statistics quality reporting.

In recent years at national and international level, it has become more and more important to care of the accessibility and clarity of the documentation about the quality of the statistics. On the other hands the credibility of the National Statistics Institutes crucially depends on their capacity to transmit trust to the users and to the public about the statistics produced. The possibility for the users to have access to information about the quality of statistics it is really a fundamental challenge for Official Statistic.

Since few years many EU Regulation require to the Member State to asses and certify the quality of their statistics with specific “Quality Report”. The purpose of those quality reports is to provide Eurostat and EU commission with tools for assessing the quality of statistics through a set of qualitative and quantitative indicators. This is very important but it is not enough: to increase transparency and to reinforce the trust of the users it would be necessary also that information about quality should be made available regularly to users and should also be easily accessible. Reports or documents on quality should be updated regularly and should include quality trends.

Assessing the quality of short term statistics is particularly difficult. This is the reason why the quality reporting of the quarterly and monthly indicators is still very rare. For example, although a discussion about quality reporting has been going on for many years in the STS working group at Eurostat, according to the STS Regulation, member state are still not obliged to produce regular quality report on short term indicator. Only few feasibility studies and preliminary tests have been carried out.

For these reasons it seems useful to circulate papers and documents on those few experience in which the quality reporting has been introduced. The LCI (Labour Cost Index) Regulation (450 /2003) require in article 8 that all Member State have to provide to Eurostat an annual quality report. The Report presented in the following pages has been written at the end of 2004 and it represents the first experience of the Istat Direction of short-term statistics (Direzione centrale delle statistiche economiche congiunturali sulle imprese, i servizi e l’occupazione). Then it is still at a very preliminary stage but sufficient to cover all aspects of quality. Also for this reason, we would welcome any comments or feedback on this report, on its structure and its weakness. For any information please contact Fabio Rapiti at this e-mail: oros-info@istat.it .

Istat transmitted the quality report to Eurostat in December 2004.

Short term statistics quality Reporting: the LCI National Quality Report 2004

Contents

0. Introduction
1. General information on national LCI and OROS indexes
2. Relevance
3. Accuracy
 - 3.1 General aspects
 - 3.2 Non-sampling errors
 - 3.3 Sampling errors and revision error
4. Timeliness and punctuality
5. Accessibility and clarity
6. Comparability
7. Coherence with related survey and national accounts
8. Completeness
9. Back data
10. Cost and burden
11. Conclusions

Introduction¹

This first quality report intends to cover the various aspects relating to the assessment of the quality in LCI statistics, as described in the art. 2 and annex 1 of R. n. 1216/2003 covering the data 1.1996-4.2003.

The structure of this report is influenced by the fact that the LCI implementation is still ongoing and only 2003-2004 data can be considered a regular delivery, while all other data must be considered back data.

The organization of the report follows largely the official criteria on statistics quality defined in the EU Regulations. The next paragraph contains a brief presentation of the main characteristics of the basic data (INPS) and the OROS survey both used to compile LCI indexes. The other paragraphs deal with the following issues:

- *relevance*: refers to the degree of having met the needs and expectations of users or user groups;
- *accuracy*: defined by the size of the possible gap between measurement and true but unknown population parameter;
- *timeliness and punctuality*: defined respectively by the time span between the reference period and the time of actual data delivery (timeliness) and the deviation of the actual time of data transmission from the target date of delivery (punctuality);
- *accessibility and clarity*: refers respectively to the conditions under which the user may obtain data (accessibility) and to explanatory information supplied in order to support comprehension and adequate interpretation of the data (clarity);
- *comparability*: gives prominence to the differences in applied concepts, definitions and methods and their effects on the interpretation of data coming from different geographical units, or different points in time;
- *coherence*: refers to the adequacy of combining survey results with data from other sources;
- *completeness*: describes the degree to which available information meets the requirements defined within the European Statistical System;
- *Cost and burden* due to the implementation of the regulation: although this issue is not included in the annex 1 of Regulation n. 1216/03, as far as we are concerned, the cost and burden on enterprises it is a very relevant aspect of quality.

This Report contains mainly a qualitative description of quality, but also some quantitative indicators which in future could constitute very useful time series. Qualitative and quantitative indicators should be user-oriented, meaning that users should have the possibility to see the extent and nature of progresses made. This version of the report has been written to fulfil the Regulation request, but in next future it could be used also to provide information to users.

1. General information on national LCI and OROS indexes

The Italian LCI is based on the OROS survey. OROS stands for *Occupazione* (Employment), *Retribuzioni* (Wages), *Oneri Sociali* (Other Labour Costs). The main aim of this survey is to produce short term information on the quarterly changes of gross wage, other labour cost and total labour cost for Italian firms in the private sector (sections C to K of the Nace Rev.1.1 classification) with at least one employee. The OROS survey, based mainly on the administrative data collected by INPS

¹ I thank C. Baldi, F. Ceccato, M. C. Congia, S. Pacini, D. Tuzi for their essential help in carrying on the Oros survey, for the processing of many indicators used in this Report and for their comments. I thank also G. Oneto and L. Tronti for useful suggestions. I remain the only responsible for any error.

(National Social Security Institute), is aimed at covering all firms size classes without increasing the statistical burden on firms. The survey has been designed also to satisfy the Short-Term Statistics (STS) Regulations.

The OROS statistics are based on different sources depending mainly on firm size: the census of large firms (more than 500 employees) data are drawn from the Monthly Survey on Labour Input and Remuneration variables in Large Firms (hereinafter Large Enterprises Survey - LES); the Small and Medium Enterprise (SME) are estimated on the base of a large non-random sample of INPS data².

From the end of 2004 Istat will deliver an actual hourly labour cost index while at the moment, thanks to a transition period for hours worked, in the indexes the denominators refer to full time equivalent. Data about “gross wages” per Fte (full-time equivalent), “other labour cost” per Fte and “total labour cost” per Fte are expressed as indexes (2000=100).

INPS submits data files to Istat three times every quarter, a non-random sample of DM10 forms, the whole population³ of DM10 forms and an update of the INPS Business Register. Each quarter two new estimations are released: the “preliminary” estimate based on a “non-random” sample of INPS data, with a delay of about 90 days from the reference quarter, and then a revision estimate, called “final”, based on the “total population” of INPS data, with a delay of 15 months from the reference quarter.

The labour cost is the sum of two components:

- gross wages, which comprises all the payments, both regular and not regular, including worker social security contributions and taxes (wages and salaries in kind are excluded);
- employer social contributions (only actual legal contributions with the exclusion of the imputed ones).

At the moment the employers’ social contributions indexes exclude the subsidies received by the employer.

It is very important to underline the similarity and differences between the OROS and the national LCI indexes. Istat has started to release regularly OROS indexes in October 2003⁴. In the quarterly OROS survey press release Istat publishes current weighed indexes at sections and aggregated levels (i.e. the total C-K at quarter t is weighted with current employment at t). The LCI annually chain-linked Laspeyres index for industry, services and total private economy (C to K) are regularly produced using OROS databases and delivered to Eurostat but not yet released at national level.

The rationale for compiling a fixed-weighted index (like the LCI) as opposed to a simple current weighed labour cost index (like OROS) is that theoretically the first one can control for changes in a number of factors, namely shifts in the⁵:

- composition of the workforce across industries (e.g. moves away from the production sector towards services);
- quality of the workforce within industries (e.g. shifts towards higher skilled occupations);
- human capital of the labour force, which might be measured by the level of educational attainment of employees.

² The OROS survey population is divided into four subpopulation:

1. the Small and Medium size Enterprise-SME (INPS data);
2. Large Firms which do not enter into the LES population (INPS data);
3. Large Firms within LES;
4. Firms of any size which offer “interim employment services company” (INPS data).

³ The population of the DM10 declarations include all businesses, also the ones already delivered in the sample.

⁴ The first occasional OROS “Statistiche in Breve” were released in November 2002, April and July 2003.

⁵ See “Review of the Labour Price Index by the United Kingdom”. Paper presented at Working Group Meeting on Wages and Labour Costs Statistics during April 2002.

The European LCI, compiled as an annually chain-linked Laspeyres index based upon a fixed structure of economic activity at NACE Rev.1 section level, addresses the first issue, controlling for shifts in the composition of the workforce among sections. Thus, the Italian LCI sections indexes remain current-weighted indexes while the aggregate index is a real fixed-weighted index. The OROS indicators instead are current-weighted labour cost indexes regardless the level of aggregation. Therefore, at section level the OROS indexes released at national level correspond exactly to the LCI indexes, while the aggregate indexes diverge.

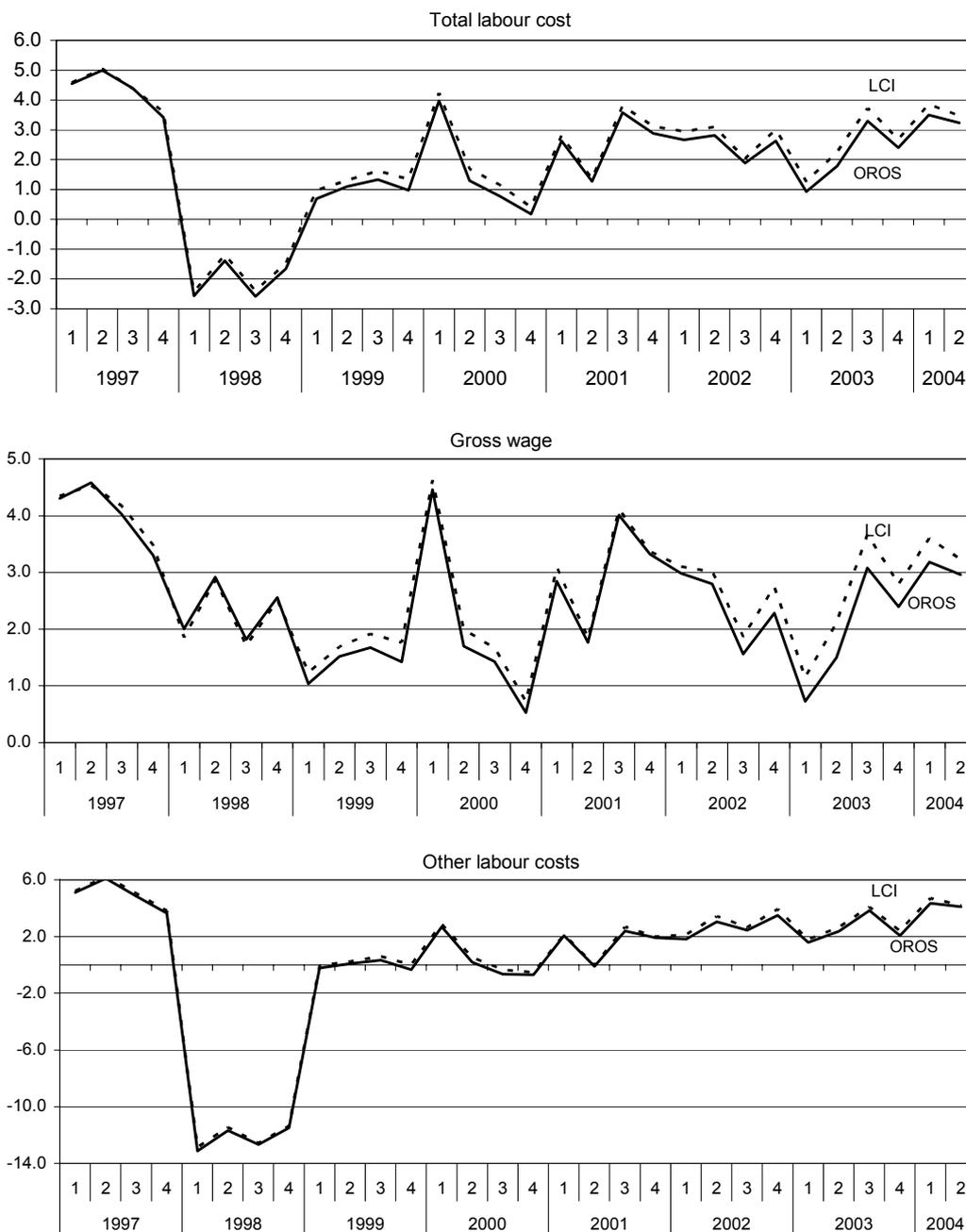
The differences in the two aggregate indexes for the total private sectors (C-K) show how the shift in employment among economic sections effect wages and labour cost.

Figure 1 highlights the higher growth of the LCI index as compared to the OROS one. The difference is the effect of the employment reallocation across industries: in the 1996-2003, and especially in 1999-2000 and 2002-2003 periods, there has been a very strong growth of employment in Construction (F), Hotels and Restaurants (H) and Real estate, renting and business activities (K), all sections characterized by very low average wages, compared to other sectors. On the contrary, there is no growth or a decline of employment in two high wage sectors, sections J (Bank and assurance) and E (Electricity, gas and water supply).

Thus, in the next paragraphs when speaking about the national aspects of the LCI we will frequently refer to the LCI-OROS data⁶.

⁶ For an in depth description of the OROS methodology see Baldi, Ceccato et al 2004.

Figure 1 - LCI (fixed-weighted) and OROS (current-weighted) indexes, quarterly year-on-year growth rates, unadjusted series I.1997-II.2004



2. Relevance

The LCI (OROS) results should provide a sound empirical foundation for decision-makers in national wage negotiations and social and economic policy, and should also establish reliable and harmonized comparisons of labour costs and its components between European countries. The main users of the survey should be:

- at the national level: the Ministry of Labour and Welfare, the Ministry of Economy, trade unions (CGIL, CISL, UIL), employers' associations (Confindustria, etc.), Member of Parliament, political parties, research centres, universities and the media.
- at the international level: the services of the European Commission, the European Parliament, ECB, OECD, IMF, ILO, etc.

Traditionally in Italy the most important indicator to monitor short-term wage and labour cost developments is the monthly Collectively Agreed Wage Index. But other indicators are needed to monitor the actual wages and labour cost trends (as opposed to agreed wages)⁷. Short term information about actual wages and labour cost was collected and released only for those enterprises in industry and services classified as “large”, that is with more than 500 employees, through the Monthly Survey on Labour Input and Remuneration variables in Large Firms (LES). For many years users were not satisfied by this situation⁸. The new LCI-OROS data, based on the administrative data collected by INPS, fill this gap and cover all firms size classes without increasing the statistical burden on firms. This is the reason why all users really welcomed the new LCI-OROS data since the first release. The new survey has become quickly another very important instrument for backing up today's social and economic decision-making.

For example, since January 2004 the Ministry of Labour and Welfare presents and comments in his bi-monthly *Nota Flash* (printed and on-line on web-site) the OROS indexes. They have been commented also in the influential Bank of Italy biannual bulletin (March 2003) and Annual Relation of the Governor (May 2004). In April 2004 for the first time the OROS indicators have been presented and discussed also in the second volume of the General Report on the Economic State of the Country 2003 (*Relazione Generale sulla situazione economica del paese-2003*) presented by the Ministry of the economy to the Parliament.

Recently, also the main trade union research institutes (IRES-CGIL, *Centro studi* CISL) have started to consider the OROS indexes among the most important indicators of the actual wage trends⁹.

Every quarter traditional media, mainly newspapers, devote many articles and quite a lot of space summarizing the press release or commenting the results of the OROS data¹⁰. An independent very influent on-line web site run by eminent economists (www.lavoce.org) has published interesting articles using OROS results. Also the OECD Italian desk has started looking at the data in their analysis.

Thus, there is enough evidence that since October 2003, when the first national regular press release has been published, main users seems interested and satisfied with the OROS data (see also § 4).

3. Accuracy

3.1 General aspects

There is always a trade-off between accuracy and data collection burden. For the LCI-OROS survey Istat has chosen to get a huge quantity of good quality administrative data, avoiding new heavy burden to business and lessening the cost for the statistical service. Obviously this has meant to face and solve traditional and new administrative data problems. Thus the evaluation of the accuracy of LCI-OROS data must take into account two main aspects of the survey:

1. they are based, mainly, on social security data (integrated with Large Enterprises Survey – LES);

⁷ An indicator used frequently is also the quarterly national account (ESA 95) compensation of employees divided by the numbers of employees expressed in fte.

⁸ See the relation of I. Cipolletta (Confindustria Director General) in the fourth National Statistical Conference (Istat 1999, p.129).

⁹ A very recent report of the IRES-CGIL underline that the OROS indexes should be taken as main wage trend indicators for wage negotiation in non agricultural private sector (IRES-CGIL 2004). Also the CISL *Centro studi* in his irregular Short Term Note (*Nota congiunturale*) publishes comments on OROS data.

¹⁰ At national level also the Eurostat quarterly LCI press release received attention from Italian newspapers.

2. the preliminary estimation is based on a non-random sample, while the revised “final” estimation is based on the total population.

In the next paragraphs we distinguish the non-sampling errors and problems which affect both preliminary and final estimations and may arise from administrative procedures or more general problems (coverage, measurement, processing and non response treatment, etc.), from the sampling error/problems which concern only the preliminary estimation. From this point of view, theoretically one of the main aspect characterising accuracy should be the calculation of the mean square error of preliminary estimates. But this is hardly applicable, because the quarterly indicators are not obtained according to standard textbook procedure, and thus the standard errors of the estimates are almost meaningless.

However given the fact that LCI-OROS indicators are released in a preliminary version followed by revised and final figures, the revisions can be seen as the realisation of a stochastic process whose properties are intrinsically linked to the accuracy of the estimates¹¹. The study of the size and direction of revisions can serve to assess the lack of any persistent or predictable biases.

The chapter is organized as following: paragraph 3.2 presents an almost exhaustive description of non-sampling errors; the following paragraph defines the revision indicator and presents some statistics on the size and volatility of the revision process.

3.2 Non-sampling errors

3.2.1 Administrative data concepts

The OROS survey is largely based on administrative social security data. Before using those data theoretical and empirical studies led Istat to the belief that all INPS variables (gross wages, other labour costs and employment) correspond to the LCI statistics concepts. From a general standpoint, some difference between the LCI statistical concepts and definitions (as stated in the Regulation) and the administrative INPS variables available, may rise mainly for two different reasons: a) the definitions are, at least partially, formally different; b) although the formal definitions are the same, there are specific (administrative) incentives for the enterprises to give incorrect information. Both problems have been analysed by comparing in details the theoretical definitions and through an empirical large micro matching experiment¹². The results show that (see § 1.7 and § 3 in Istat 2000):

1. for all variables the difference in level between the LCI definitions and the INPS variable may be considered extremely limited;
2. in practice, many enterprises, when responding to the direct statistical surveys, enter on the Istat questionnaires the same variable already reported on the INPS form;
3. even with little possible differences in levels, these differences do not necessarily show up as different trends in the wage change.

Moreover the conversion of the administrative data into the required statistical variables implies computational aspects. The INPS data go through a very complex process: the first stage consists in the retrieval of the statistical variables using trans-coding and aggregation of the elementary variables present on the raw data. In this phase, it is necessary to re-aggregate several “employment” type and “contribution” type variables associated to codes present on the DM10 on the basis of the contribution

¹¹ “Presentation of the quality report for the PEEIs”, Eurostat STS working group documents, 30 September, 2004.

¹² The microdata matching tried to compare and evaluates the quality of the INPS data as compared with another business statistics survey variables, used as benchmark. For the latter, an internal Istat statistical source was used. Univariate and multivariate analyses were used for measuring the size of the distortion and for evaluating its systematic or casual nature. The results obtained show how the majority of the problems connected with the content of the INPS variables turn out to be, in practice, of a very limited extent (Istat 2000).

homogeneity. In this phase could arise (metadata) problems in identifying all little administrative labour cost items and sub items (characterized by codes and sub codes: more than 800) and this could sometimes bring to undetectable basic coding errors (wrong reconstruction of wage and other labour cost variables) especially in reference to “other labour costs”¹³.

3.2.2 Register, coverage and misclassification errors

The OROS Business Register (O-BR) derive from the combination of the official Business Register (called ASIA)¹⁴ maintained by the Business Register Division of Istat and from a list of all businesses registered in INPS’s Administrative Business Register (A-BR)¹⁵.

INPS updates regularly the A-BR. Changes are essentially related to the registration of new units and the change of information over units which already exist. One of the main negative characteristic of the A-BR deals with the quality of the activity status (or the lack of it). While all births are registered because it is mandatory to enrol to INPS in order to set up an enterprise with employees, the firms deaths are not frequently removed from the AR because, although it should be compulsory, there is no administrative incentive to enforce it (penalty for the firms or interest of INPS personnel to register it). This is one of the main source of error in the SME population in the preliminary estimation: over coverage of the target population.

The process of construction of O-BR starts from the capture of the A-BR, which is made available to Istat at the end of each reference quarter and contains information for about 2 million administrative units. Then it undergoes some phases of checking and it is matched with the ASIA BR, mainly to acquire the economic activity code (which is of much better quality as compared to the A-BR one). The classification rule consists in drawing the NACE code from ASIA where the two register match and in using the NACE code of the A-BR for the residual units. About 70% of the units in the register get the economic classification from the ASIA Business Register¹⁶. Thus, misclassification may arise sometimes from the unit with the Nace code assigned from the A-BR. It is useful to remember that for large companies classification errors are almost impossible because the data derive from the LES survey.

Over-coverage errors, mainly trough duplications of unit or “part” or firms may result from missing or broken links between corresponding units in the two different sources which cover the whole population in a complementary way (SME deriving from INPS data and LES). For example after a business restructuring in which a new fiscal code number, identifying the legal entity, is adopted but it has not been yet up-to-date in one of the two sources and, a duplication of unit may result. Mergers or company splits may also give rise to similar over-coverage errors. Very seldom also under-coverage could result from missing or erroneous fiscal codes or legal information in the A-BR.

For large companies both coverage errors and duplications are identified during a specific phase of checks in which LES and INPS data are carefully analysed and compared. If SME’s are involved, such errors may remain undetected. These errors are the hardest to detect but they should also have an irrelevant impact on the estimate.

3.2.3 Economic activity, type of employment and labour cost items coverage

¹³ This may happen because the input data change continuously: not only a continuous change of the data but also of the metadata, i.e. social security codes and meaning of the codes themselves.

¹⁴ The Business Register ASIA (*Archivio Statistico delle Imprese Attive*) is a list of all businesses in Italy; it is updated each year using data from various surveys and administrative data (Tax Register, Register of Enterprises and Local Units provided by the Chambers of Commerce, Social Security Register, Work Accident Insurance Register, Register of the Electric Power Board) treated with statistical methodologies.

¹⁵ The original A-BR reporting unit is the “administrative social security unit”, which may or may not correspond to a legal enterprise. By linking all the “administrative social security positions” through their fiscal code it is possible to reconstruct the firm.

¹⁶ The reason for the non-correspondence between the two archives is mainly due to the temporary release gap: the ASIA Business Register is available with up to two years delay from the reference quarter of INPS A-BR.

All economic activity between sections C to K (Nace rev. 1.1) indicated in the Regulations are covered.

The indexes are calculated including the following three categories of employees: manual workers, non manual workers and apprentices; they do not include managers. The coverage of those three categories in the two sources (LES and INPS) data is complete. The employees represented by the sum of the two sources compared to the number of employees according to ESA 95 are presented in the Table 1.

Thus the managers are the only missing group of employees excluded from the coverage and this means that the coverage of LCI-OROS indexes is more than 98% (9,516,020/9,704,400). There is no model or estimation criterion used to adjust the data to include managers. The impact of this exclusion has never been calculated.

All labour cost items are collected at monthly frequency from the same two sources. Only the variable “subsidies received by the employer”, at the moment excluded in the calculation for the transition period, will be drawn from administrative annual information.

Table 1. - Coverage of the LCI-OROS survey compared to NA ESA-95 (number of employees in C to K nace sections), 2001

	LCI-OROS (1)	National account annual data, ESA-95 only regular employment (2)	(1)/(2)
Manual, non manual and apprentices	9,516,020	-	-
Managers ^(a)	122,095	-	-
Total	9,638,115	9,704,400	99,3

(a) From the OROS survey it is possible to estimate the number of mangers but not their wages.

3.2.4 Delineation of Unit

The observational units in the LCI-OROS data are, to a large extent, business units rather than (the more ideal) kind-of-activity units (KAU). A firm which is classified as being part of NACE group G (because of its main activity) may produce some products or services which are associated with another NACE group, say J. Consequently, this unit should ideally contribute to the index for two NACE groups G and J. However, in the estimation of the index this possible mix is ignored and all output of the same firm contributes to its main NACE group. This approximation has been judged acceptable.

3.2.5 Hours worked

The actual indices are calculate per FTE (Full-time equivalents). Istat will estimate and deliver the new LCI hourly indices in December 2004. For a first detailed description of the methods for compiling the hours worked see Istat 2004. Here it is worth noting that for producing quarterly estimates of per capita hours worked for the period 1996-2004 it is necessary to use different sources and an indirect estimating procedure (more exactly, benchmarking). By benchmarking procedures, we mean those models in which annual data are broken down over shorter time periods (quarters) with the help of one (or more) reference indicator(s). The annual information derive from SBS statistics and among the short term indicators we will use LES and LFS data. Once the per capita quarterly hours worked have been estimated, the total number of hour worked is calculated by multiplying the quarterly per capita

hours and the “number of employees” (estimated within the OROS survey and coherent with the numerator). By dividing the total wage bills and the total other labour costs (estimated both with the OROS survey) by the total number of hours, the labour costs per hour worked is obtained.

3.2.6 Non-response adjustment

Non response adjustment is used only in the final estimate and only for the SME subpopulation. The main problem is the distinction between a normal absence of a DM10 (seasonal activity, death, etc.) and a real unit non-response. In order to distinguish the two very different events, auxiliary information on the units are used. To avoid over imputation, the units which belong to small size classes (less than 20 employees) have been excluded from imputation. The imputation method uses both longitudinal micro and cross section cells information. In terms of total employment, the imputation implies an increase on the number of employees amounting to less than 2% (on a share of units to be imputed which is about 1%).

3.2.7 Processing errors

There are two types of processing errors:

1. system and programming errors (‘bugs’)
2. data handling errors.

It is very difficult or even impossible to assess the impact of these sources since there are no data available that could form a basis for an analysis of the data processing operations. However, some quite general comments could be made.

System or programming errors are a serious potential error source. The LCI-OROS procedures imply long and very complex SAS programs. These errors are minimised by checking and testing procedures during the design of the system. Programming errors with a very large impact on the results are likely to be detected. But programming bugs that cause results that are wrong but still appear to be reasonable could be undetected for a long time. Unfortunately, at the moment, we cannot estimate this effect in the LCI-OROS.

There are various types of data handling errors. The main sources in the LCI-OROS survey are the data transmission and data editing.

Data transmission from INPS to Istat and from Istat central server to the OROS operative server theoretically can generate errors. There are many checks which should guarantee no errors.

The purpose of editing is to detect and correct errors but it is known that editing could also sometimes introduce new errors that were not there in the first place. From a general point of view, the OROS editing and imputation procedures are quite light: the procedures are designed not to over edit or over impute. Eventually large data entry errors should be easily discovered and corrected. For this reason it is very unlikely that data entry errors are a major error source.

3.2.8 Check and editing

Once the DM10 forms are captured and all variable, are reconstructed, extensive edit procedures check for accuracy and consistency. Given the enormous number of records to check, the editing procedures are very selective. The automatic editing is reserved for very few cases. The micro editing process include the comparison of the values of all variable available (e.g. number of employees, wages per employee, other labour costs, etc.) reported for two consecutive months by the same unit to detect large changes or errors in reporting. Whenever there is a very large change or the data appear inconsistent in any of the variables the records are flagged: ‘3’=‘apparently hard errors’, ‘2’=‘apparently soft errors’. All the other records results with no errors=‘1’. Than the records flagged 3 or 2 in at least one variable are ordered for size of wages and other labour costs changes. The record are displayed and checked on the screen which shows a very complete dashboard with all information about the unit under control which allow the editor to understand if the anomalies are errors or just

outliers These can be acknowledged and values are further treated as correct. Then the errors are manually corrected if necessary. In principle this ought to allow to make reliable statistics, however some errors in small size units remain undetected and some could show up only after micro-data aggregation. After the calibration process another selective editing check is carried on.

3.2.9 Model assumption errors

Few data, used in some steps of the OROS survey process, are collected on a less frequent basis (yearly or once every second or third year). The application of these data to current quarters implies important model assumptions of time invariance.

A number of sources have been considered, they include:

1. Use of “average annual contractual hours of work” of the previous year in the compilation of full-time equivalent units.
2. Use of constant rate of “contribution paid by the employee” in calculating quarterly “other labour cost”;
3. In the preliminary index the methodology used in the estimation of the current population at time t (in which there is over coverage) implies that the unit in t has a probability of been active calculated with information referring to time $t-4$.

3.3 Sampling errors and revision errors

In the “final” estimation there is no sampling error since the data represent the census although there is some partial non-response (see. § 3.2.6); while in the preliminary estimate there is a sort of “sampling error” relative only to the SME subpopulation¹⁷. The sample used in the estimation, although it is non-random, is huge and growing over time, as firms adopt the electronic way to send INPS the DM10 forms. Besides, it has a good degree of coverage of the target population as regards the breakdown by economic activity sectors and the age of firms: specifically, the sample includes a large number of births.

The methodology used to obtain the preliminary estimate is built to cope with the non-random sample and to produce coherent estimates of the target variables: in fact the main parameters consist of ratios of target variables (e.g. wage per full-time equivalent employee), so the consistency of the numerator and the denominator has to be granted¹⁸. For an in depth description of the methodology see Baldi, Ceccato et al. (2004).

3.3.1 Total revision errors¹⁹

In general terms, we define “revision” as latest minus preliminary estimations and “revision error” the difference between the preliminary estimation of the indexes and the final version. The nature of this error is comprehensive of a number of factors which can be summarized in just two:

1. The most important refers to the type of estimation: while the preliminary estimate is based on a sample, albeit of big size, the final estimate, based on the universe, can be defined a census. For this reason this revision is different in nature from all the revisions implied by those surveys that release a fast, preliminary estimate based on subset of respondents that have already been

¹⁷ Another different sort of “sampling” or “imputation” error exists also for the other two subpopulations: large firms which do not enter into the LES population (INPS data); firms of any size which offer “interim employment services company” (INPS data).

¹⁸ The methodology is based on giving a weight to each unit of the sample. The weights are calculated to satisfy the condition that the sum over the units of the sample of an auxiliary variable multiplied by the weight of the unit, is equal to the known total of the auxiliary variable (calibration).

¹⁹ This paragraph rely heavily on paragraph 3.1 of the paper by Baldi et al (2005). I thanks C. Baldi and the other authors for letting me use their previous work.

contacted, and later release figures based on larger sample. Even if this general case looks similar to OROS revision, it is very different. The number of the units on which the final OROS estimate is based is far larger than those of the sample. The universe, whose size is stable at 1.3 million of units, is between 2.0-2.5 times the sample used in the estimation of 2002 and 2003;

2. The other factor refers to all other non-sampling errors (non-response errors, measurement errors, as well as processing, coding and imputation errors). Taking also in account that for administrative reasons²⁰, a relatively little but not irrelevant number of social security declarations already present in the sample are substituted and revised directly by INPS with new and better data which are delivered for a second time to Istat within the population. This implies that for the same observed units the data in the sample and in the census may be very different.

For these reasons in order to distinguish the standard revision error from the OROS error, which embodies all the difference between the survey sample estimate and the real census value, we prefer to speak about Total Revision Error (TRE).

The revision error may be used to study and understand the causes of the bias with the aim of reducing it. The study of the volatility of the revision process can provide a very precise measure of the accuracy of the estimates. Revision analysis provides thus a key tool for assessing the trade off between timeliness and accuracy.

The error in quarter t is defined as:

$$e_t = \frac{\hat{y}_t^p - \hat{y}_t^f}{\hat{y}_t^f} * 100 \quad [1]$$

where:

\hat{y}_t^p is the preliminary estimate (with prime p) of the target variable at quarter t, while \hat{y}_t^f is the final estimate (with prime f) at quarter t.

The error in the quarterly year on year change is:

$${}_{vt}e_t = {}_{vt}\hat{y}_t^p - {}_{vt}\hat{y}_t^f \quad [2]$$

where:

$${}_{vt}\hat{y}_t^p = \frac{\hat{y}_t^p - \hat{y}_{t-4}^p}{\hat{y}_{t-4}^p} * 100 \quad [3]$$

is the quarterly year on year change for the preliminary estimate and

$${}_{vt}\hat{y}_t^f = \frac{\hat{y}_t^f - \hat{y}_{t-4}^f}{\hat{y}_{t-4}^f} * 100 \quad [4]$$

is the quarterly year on year change for the final estimate.

Then the following statistics can be computed:

²⁰ The sample DM10 declarations are not checked by INPS. The file of the population instead is composed of forms which have undergone some formal and substantial checks and corrections and also the "Vigilanza" procedure in the INPS local offices. Those checks and procedures may change the contents of the economic variables.

$$MPE(vt) = \frac{\sum_{t=1}^T vt e_t}{T} \quad [5]$$

$$MAPE(vt) = \frac{\sum_{t=1}^T |vt e_t|}{T} \quad [6]$$

The first one is the Mean Percentage Error (MPE) and measures the average error *tout court*; the second is the Mean Absolute Percentage Error (MAPE) and sums up negative and positive values. A cross reading of the two indicators shows systematic bias: when MPE and MAPE are equal it means that the systematic bias is positive; if the indicators are equal but with different signs, the systematic bias is negative.

In the next table the Mean absolute and percentage revision errors are computed over revisions of the last four quarters in which the final (i.e revised) data are available.

Table 2 - TRE (Total Revision Errors) of the quarterly year-on-year growth rates, unadjusted series III.2002-II.2003

	Economic activity sectors									
	Industry and services (C to K)	Industry			Services					
		Industry and buildings (C-F)	Industry (C, D, E)	buildings F	Services (G-K)	G	H	I	J	K
Gross wages										
3-2002	-1.1	-1.0	-0.8	-1.1	-1.1	-1.3	-1.1	-1.3	-0.4	-0.6
4-2002	-0.4	-0.6	-0.5	-0.5	0.1	-0.4	-0.2	0.0	-0.3	0.8
1-2003	0.1	0.0	0.0	0.7	0.1	-0.2	-0.1	0.6	-1.4	1.5
2-2003	0.7	0.3	0.3	0.5	1.0	1.0	0.2	0.3	-0.9	2.8
MPE	-0.2	-0.3	-0.3	-0.1	0.0	-0.2	-0.3	-0.1	-0.8	1.1
MAPE	0.6	0.5	0.4	0.7	0.6	0.7	0.4	0.6	0.8	1.4
Other labour costs										
3-2002	-0.9	-0.9	-0.8	-1.6	-1.2	-1.8	-1.8	-0.8	0.3	-1.1
4-2002	0.4	-0.1	0.1	-0.7	1.0	0.4	0.8	1.0	0.6	1.4
1-2003	-0.2	-0.4	-0.2	-1.6	-0.1	-0.6	-0.6	0.8	-1.4	1.4
2-2003	0.5	0.2	0.4	-1.3	1.2	0.9	-0.1	0.7	-0.6	3.1
MPE	-0.1	-0.3	-0.1	-1.3	0.2	-0.3	-0.4	0.4	-0.3	1.2
MAPE	0.5	0.4	0.4	1.3	0.9	0.9	0.8	0.8	0.7	1.8
Labour costs (LCI)										
3-2002	-1.1	-1.0	-0.8	-1.1	-1.1	-1.4	-1.2	-1.2	-0.4	-0.6
4-2002	-0.1	-0.4	-0.2	-0.5	0.3	-0.3	0.1	0.2	-0.1	0.9
1-2003	0.0	-0.1	-0.1	0.0	0.0	-0.2	-0.4	0.6	-1.5	1.5
2-2003	0.6	0.3	0.3	-0.2	1.1	1.1	0.1	0.4	-0.9	2.8
MPE	-0.2	-0.3	-0.2	-0.5	0.1	-0.2	-0.4	0.0	-0.7	1.2
MAPE	0.5	0.5	0.4	0.5	0.6	0.8	0.5	0.6	0.7	1.5

The time path of the TRE shows that it is not constant. In III.2002 it was high. In the other quarters it has been definitely lower. The TRE is higher in sections J and K. In general, the size of the MPE is

acceptable, although especially the “other labour costs” variable in some sections (F and K) it is quite large. The comparison of MPE and MAPE shows a systematic bias in sections F and K of the “other labour costs”.

3.3.3 Revision history

As seen previously, a regular revision of the indexes is released after 15 months from the reference quarter. Just another revision has to be taken into accounts. In march 2004 the methodology to link the old series (1996=100) and the new one (2000=100) has been changed slightly. Previously the link was based on the first quarter of 2000. In the revised series the link is the average of the four quarters of 2000. The revisions had some effect on all quarters of 2000.

4. Timeliness and punctuality

According to the Commission Regulation (EC) No 1216/2003 of 7 July 2003, Member States shall transmit the quarterly results of their LCI statistics to the Commission (Eurostat) no later than 70 days after the end of the reference period. Istat had a transition period for timeliness of one year (90 days instead of 70), but it has expired in summer 2004.

The LCI punctuality is the result of a number of factors, to be assigned partly to the general policy of the NSI, to the level of technical and organizational co-operation with INPS, to the support and the quality level of services offered by different support departments (ICT) and, finally, to the wage and labour cost statistics department itself.

The OROS-LCI production process is very complex and there are still some technical and administrative problem in the phase of data transmission from INPS to Istat. At the moment the whole process still takes longer than the Regulation deadline.

An internal program for reducing the timing of the whole process is still on-going and it will allow Istat to fulfil the target timeliness before the end of March 2005.

Table 3 - National Publication Calendar and delay in the delivery of LCI data to Eurostat

Reference quarter	National OROS-LCI press release	Delay in the delivery of LCI to Eurostat (in days)
Q2.2003	16.10.2003	160
Q3.2003	28.12.2003	100
Q4.2003	29.03.2004	90
Q1.2004	28.06.2004	88
Q2.2004	29.09.2004	89
Q3.2004	22.12.2004	80
Q4.2004*	22.03.2005*	75*
Q1.2005*	17.06.2005*	70*

(*) Scheduled date.

5. Accessibility and clarity

Statistical information has the higher value when it is easily accessible by all users under equal conditions and is available in the formats that they require. The OROS results are currently

disseminated via the Istat website as well as via traditional print publications. The media are informed about quarterly main results during a briefing at Istat when a press release is disseminated. In the same morning the press release and the data are put on the Istat web site and the OROS data in the short term time series database (called CONISTAT) are updated with the last quarter.

The state of dissemination is summarized in the Table 4

Table 4 – State of dissemination of the OROS-LCI survey

Type	Frequency	Media	Main users
Press release (<i>Comunicato Stampa</i>)	quarterly	Istat web site	mass media
Conistat – time series database (www.istat.it)	quarterly	Istat web site	specialised users
Labour and wages annual publication (<i>Annuario Lavoro e Retribuzioni</i>)	annual	Print publication	all users
Istat Annual Report (<i>Rapporto Annuale Istat</i>)	annual	Istat web site Print publication	all users
Italian annual statistics (<i>Annuario Statistico Italiano</i>)	annual	Istat web site Print publication	all users

The clarity of LCI-OROS results depends on the availability of assistance provided in using and interpreting the data. This assistance is given in form of methodological comments in CONISTAT database (which uses the SDDS standard for metadata), or at the end of the press release (“*Note informative*”) and other publications. Further assistance is available by telephone or via mail at oros-info@istat.it by contacting directly the Istat unit responsible for the OROS data.

The metadata notes include information on the following:

- Concepts, Definitions and Classifications;
- Population coverage;
- Data collection methods and nature of administrative basic data;
- Seasonal adjustment;
- Glossary of main definitions.

One very gross measure of accessibility could be the number of downloads of the CONISTAT database. From July 2003 to September 2004 the number of accesses have been 7,179. This indicator cannot be used as an absolute measure of accessibility but rather as a measure of evolution of the accessibility when, in the next future, we will have historical series of accesses.

6. Comparability

The procedures to estimate LCI-OROS data have some differences according to different time periods: there is a little break in time series in Q1.2000. Data from Q1.1996 to Q4.1999 have been estimated using only INPS data (LES data for that periods were not suitable). From Q1.2000 to the current quarter data have been produced integrating INPS with LES data. The two series have been adjusted to produce a consistent measure of trends from Q1.1996 to the current quarter. The time series incorporate also some adjustments reflecting little improvements in the calculation of other labour

costs (inclusion of social security codes previously disregarded) and in the estimation of the sub populations (interim job service firms).

Those little changes should have practically no impact on continuity.

7. Coherence with related surveys and National Accounts

LCI-OROS survey estimates may be compared with the aggregate figures coming from other sources. Such a comparison may reveal the existence of a bias from measurement error, although it may be difficult to determine to what extent the differences between estimates are attributable to errors in the LCI-OROS survey or in the other data source.

Information on wage and labour cost short term trends can be found in:

- the quarterly ESA 95 National Accounts (NA) compensation of employees per fte;
- the monthly Collective Agreed Wage Index (*Indici delle retribuzioni contrattuali*);

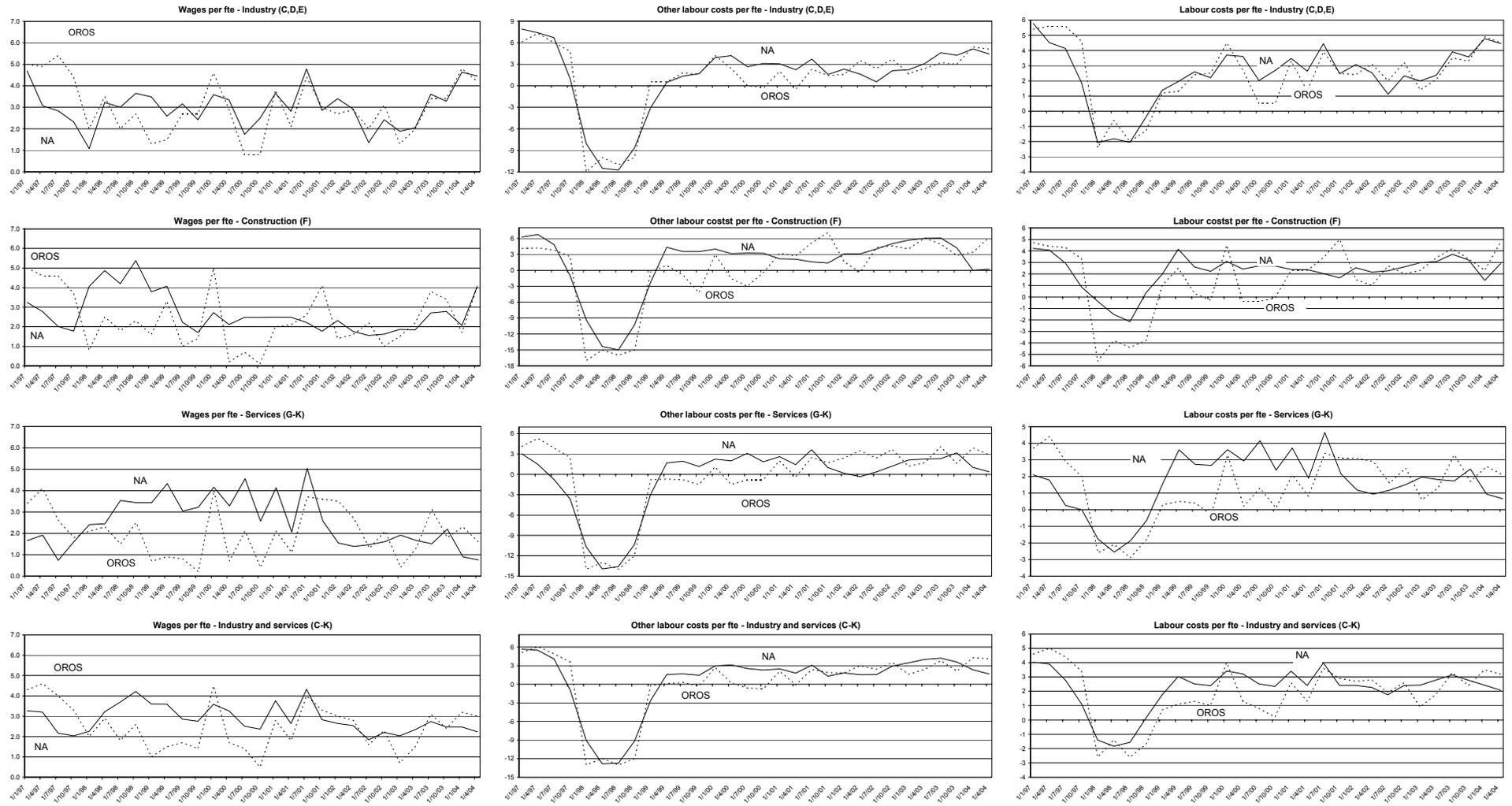
The last source refers only to the collective agreed part of the wage that, although it is very large (75 to 95 per cent of the total wage, according to different economic sectors), do not represent the actual gross wage. Thus this paragraph is devoted to the comparison of NA and LCI-OROS.

Three groups of graphs (figure 2) show the time series (1997-2004) of the annual unadjusted growth rates: wages per fte, other labour costs per fte and total labour cost indexes for NACE Rev. 1 sections and NA ESA 95 gross wages per fte, social contributions per fte and total compensation of employees per fte.

From a general point of view the pattern are quite similar, especially in the last period: the differences in the growth rates are very little for the last 12 quarters.

In industry (C,D,E) the LCI-OROS and NA show very similar trends both in wages and other labour costs, and consequently in total labour costs.

Figure 2 - Annual unadjusted growth rates of LCI and NA ESA 95 variables (Q1.1997-Q2.2004)



In services (G to K) the indicators in some periods, especially from Q3.98 to Q1.01, shows quite different pattern. There is no simple explanations for those differences.

When comparing LCI-OROS indexes and quarterly NA data it is worth not forgetting the large differences in coverage, concepts, definitions and classifications. The Quarterly NA data coverage in term of employment include regular and irregular workers (those latter workers active in the underground economy have a share of about 16% on the total work); irregular workers are obviously absent in the LCI-OROS survey. The Quarterly NA data include managers, which are excluded in the OROS survey. The Quarterly NA data are estimated on accrual basis, following the ESA 95; as opposite OROS data refer to cash. Contrary to the NA, the OROS uses the enterprise as the statistical unit, not the local Kind of activity unit. The variables have some differences in definitions: in NA the gross wages include the wage in kind and tips. OROS do not cover tips and includes only very partially wages in kind.

8. Completeness

The LCI-OROS survey went through an initial phase, when derogations existed in most member states. For Italy these derogations referred to exceptions to the standards and procedures agreed on:

- the 70 days transmission delays that were temporarily extended up to 90 days;
- labour costs per fte in the place of hour worked for current and back data;
- employers' social contributions plus taxes paid by the employer less subsidies received without treatment of taxes and subsidies (D4 and D5);

This is the timetable of the future implementation of Regulation (EC) No 450/2003:

- the LCI per hours worked will be produced and delivered in December 2004;
- the 70 days transmission will be reached in march 2005;
- the inclusion in the LCI definition of the “subsidies” will be completed in June 2005.

9. Back data

All back data, in the OROS survey context, are obviously, “final” estimations. From a general point of view the back data have been produced with the same methodology employed for current final estimation, but with some exemptions. Between back data and the current data there is no difference in terms of the coverage of economic activities, employees, labour cost items. One main difference refers only to the data from 1996 to 1999 (see § 6).

10. Cost and burden

The LCI Regulation takes into account the importance of the increasing burden on enterprises²¹. But this quality component, and especially the cost and burden on enterprises due to the implementation of the regulation, is not explicitly highlighted in the annex of the commission regulation where quality characteristics are listed.

To produce LCI-OROS data Istat did not increase at all the burden on enterprise because it was used a pre-existents survey (LES-Large firms Survey) and administrative data. The number of enterprises surveyed by the LES is less then 1000, and no additional information was requested for the LCI objectives.

²¹ In the introduction whereas of the Regulation “The benefits of collecting, at Community level, complete data on all segments of the economy should be balanced against the reporting possibilities and the response burden on small and medium-sized enterprises (SMEs)”.

To get to the same results with a traditional business survey, more than 15,000 firms should have been surveyed and as a result a new heavy burden on business and high costs for NSI would have emerged.

At the moment Istat has access free of charge to the administrative data thanks to a general framework agreement with INPS. This reduces very much the current cost of production of the survey. The main cost item is the human resources employed, which in the last three years were an average of 6 full-time employees.

The OROS survey has been the first national experience in producing short term current statistics with administrative data. Istat had to face very new and complex social security data handling problems. Thus, in the very first phase of the project Istat had to meet some additional costs when a very new type of survey had to be experimented and implemented.

11. Conclusions

The design and implementation of the LCI-OROS survey has taken few years. The next and final phase of implementation will last until June 2005 when all the obligations will be met. The Quality Report shows that the LCI-OROS survey succeeds in reconciling some contrasting needs: accuracy, coherence, comparability, timeliness, zero new burden on business and relatively low cost for NSI. Although at the cost of some difficulties and delays the design and establishment of the new OROS survey and the current production of LCI indexes can be considered a successful experience.

Reference

Baldi C., Ceccato F., Congia M. C., Cimino E., Pacini S., Rapiti F., Tuzi D. 2004; “Use of Administrative Data for Short Term Statistics on Employment, Wages and Labour Cost”, in Essays Istat (Proceedings of the “17th Roundtable on Business Survey frames) Rome, 26-31 October 2003

Baldi C., Ceccato F., Pacini S., Tuzi D. 2005 “La stima anticipata OROS sull’occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento”, forthcoming in Contributi Istat 2005

Eurostat, 2002, “Review of the Labour Price Index by the United Kingdom”. Paper presented at Working Group Meeting on Wages and Labour Costs Statistics during April 2002.

Eurostat, 2004, “Presentation of the quality report for the PEEIs”, STS working group documents, 30 September,.

Istat, 1999; “Intervento di I. Cipolletta, Direttore Generale di Confindustria”, in Atti della Quarta Conferenza Nazionale di Statistica 11-13 Novembre 1998, Tomo I- sessioni Plenarie, Istat.

Istat, 2000, “Use of INPS administrative data for the production of labour cost index (LCI) in Italy”, Intermediate report to Eurostat, September 2000.

Istat 2004, “Measuring and estimating the hours actually worked, in the contest of the labour cost index – LCI.”, Report on the implementation of LCI to Eurostat, March.

Ires CGIL 2004, Salari, inflazione e produttività in Italia e in Europa, Report presented 8th September.