Breaks in the LFS and their treatment: the Italian experience

on survey changes in 1992 and 2004, on the transition to new NACE and ISCO, on changes of boundaries in NUTS regions.

Antonio R. Discenza   *discenza@istat.it*

Francesca Gallo   *gallo@istat.it*

Alessandro Martini   *alemartini@istat.it*

Simona Rosati   *sirosati@istat.it*

Andrea Spizzichino   *spizzich@istat.it*

ISTAT – Italian National Statistical Institute

Labour Force Survey Division

Objective

Overview of main changes occurred in the Italian LFS in the last two decades and the solutions adopted to disseminate consistent time series and databases.

Outline of the presentation

- 2004: introduction of the continuous survey

- 1992: change of definitions, sample stratification and weighting procedure

- 2008: transition to the new NACE Rev2 classification

- 2011: transition to the new ISCO08 classification

- 2012: introduction of 3 new NUTS3 and modification of two NUTS2 and NUTS1 regions.

Istat

The change from "quarterly" (QLFS) to "continuous" (QLFS) in the first quarter 2004

The most complex and greater change in the LFS history.

LFS was completely renewed:

 ⤵ to fulfill Eurostat Regulations,

 ⤵ to improve reliability and quality,

 ⤵ to widen the contents for national users.

Major changes and improvements were made also to the methodology, organization, survey design, data collection strategy, interviewing techniques.

How to back re-calculate the time series from 1992Q4 to 2003Q4 ?

Istat

Project development for new LFS

When the project for the new LFS started it was already clear that:

- the effect of all these innovations would have introduced not negligible breaks on the time series, even for the main indicators

- the new LFS data would have highlighted a seasonal pattern quite different from the old one;

- it would have made impossible to compare the new data with those disseminated up to the last quarter of 2003, creating serious problems for the long- and short-term analysis, and for the production of seasonal adjusted series;

- all these changes would have made impossible to apply a micro approach for back-recalculation of old time series.

Istat

The strategy used to allow back recalculation of time series

Was based on the following points:

- continue to carry on the old QLFS and disseminate their estimates up to the first quarter 2004;

- start to observe the labour market with the new CLFS since the first quarter 2003, and start to produce the new estimates for internal use;

- fit statistical models needed for back recalculation, using the 5 overlapping quarters, taking into account the seasonal effects

…. from 1959	Q 1 2003	Q 2 2003	Q 3 2003	Q 4 2003	Q 1 2004	Q 2 2004

QUARTERLY

CONTINUOUS
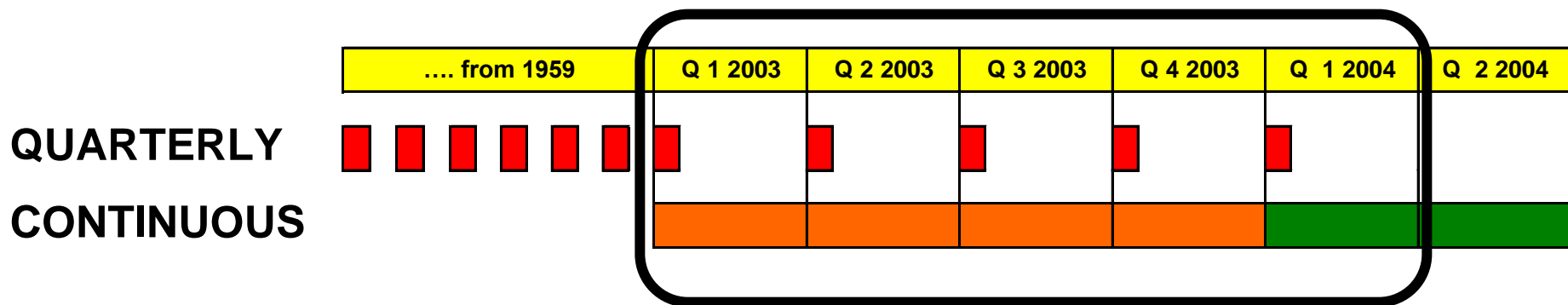
Methodology applied for back recalculation of quarterly series 1992Q4 – 2003Q4

Is based on three main principles:

✓ macro founded:
huge amount of innovations does not allow a micro-founded approach

✓ decomposition based:
consists of decomposing the series in three components: cycle-trend, seasonal, irregular, performing separated back recalculation of the elementary components and then aggregating them

✓model based
based on the hypothesis that the temporal components of the CLFS series are a function of the corresponding components of the QLFS series

Istat

The Main Hypothesis

thus concern the relationship between the series in the overlapping period:

➢ The CLFS series are function of the QLFS series

$$Y_t^{CLFS} = f(Y_t^{QLFS})$$

➢ It is possible to decompose the two series trough traditional time series analysis methods, with an additive model for seasonal decomposition, where the cycle-trend term include the irregular term.

➢ The relationship between the series is still valid at the elementary components level, hence

$$Y_t^{CLFS} = f_1(T_t^{QLFS}) + f_2(S_t^{QLFS})$$

Dissemination of results

✢ Back recalculation of a huge amount of quarterly series was successfully completed during the first half of 2004

✢ results were published at the same time with the first press release of the new continuous LFS (2004Q1)

dissemination time series of the main characteristics (employment status, 11 groups of economic activity, professional status, job duration, full-time part-time, duration of job search, etc) broken down by sex and five-year age classes at NUTS2 level.

✢ seasonal adjusted data of the overall period 1992Q4-2004Q1 were also disseminated

Since 2004 major efforts have been focused on improving the detail of disseminated series.

Istat

Employment 1992 - 2004
(absolute values in thousands)
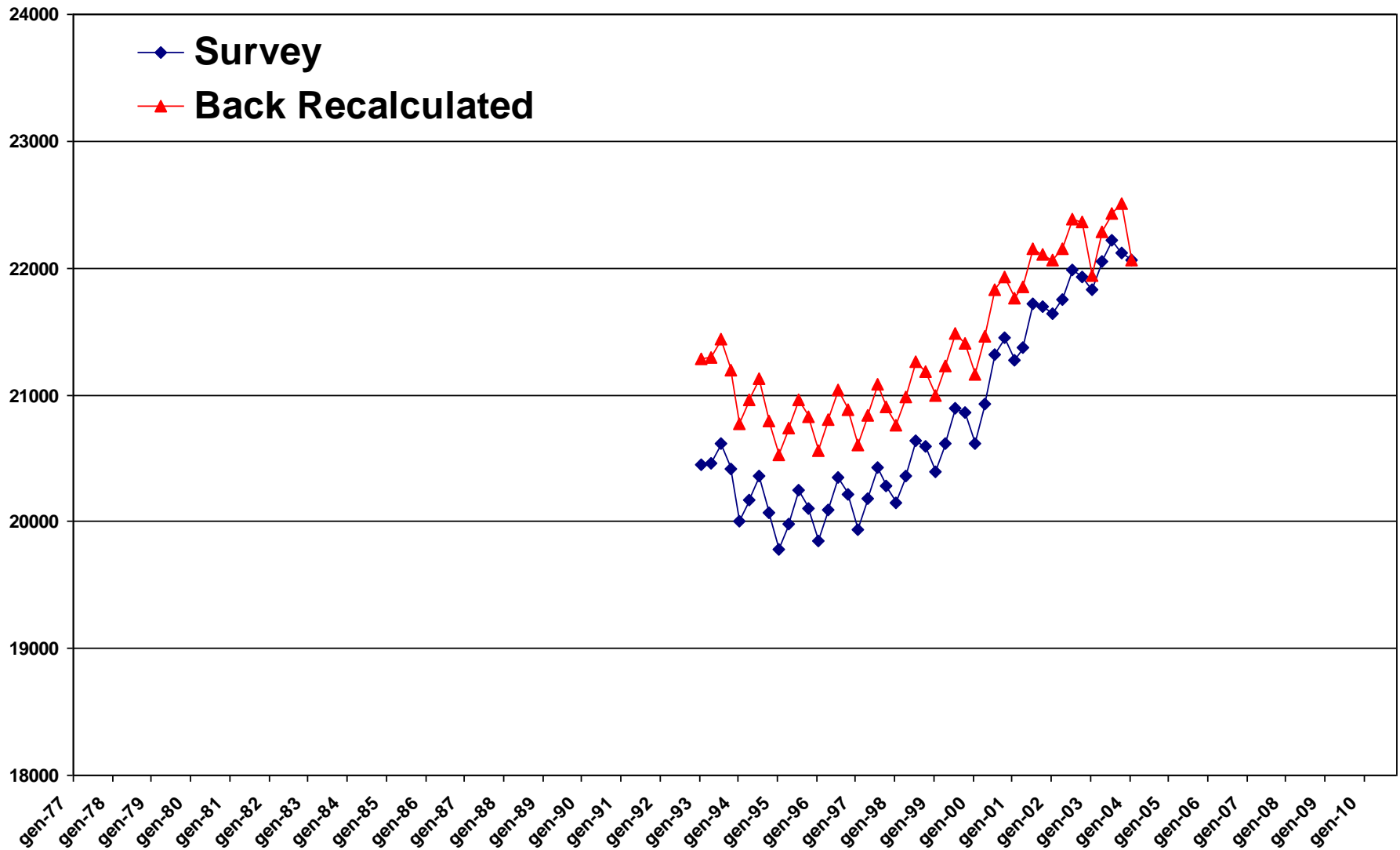
Back recalculation of time series 1977 – 1992 consistent with the recalculation 1992 – 2003

- In the last few years LFS unit wanted extend the production of labour market time series to the period before 1992.

- In April 2013, a new back recalculation for the period 1977Q1-1992Q3 was thus added to the current series 1992Q4-2012Q4, obtaining a set of time series covering the last 35 years.

- These series are perfectly consistent with each other, at all levels of aggregation. Future work is already planned in order to increase the details of information.

- Are obtained starting from several sets of time series that were already recalculated in the past in different context, and for different purposes.

The framework for
back recalculation procedure

↳ **Three branches** of different "original" survey data were available

↳ **A two-step procedure** was applied for the period **1977Q1-1992Q3**

CLFS data
2003Q1
2012Q4

QLFS data
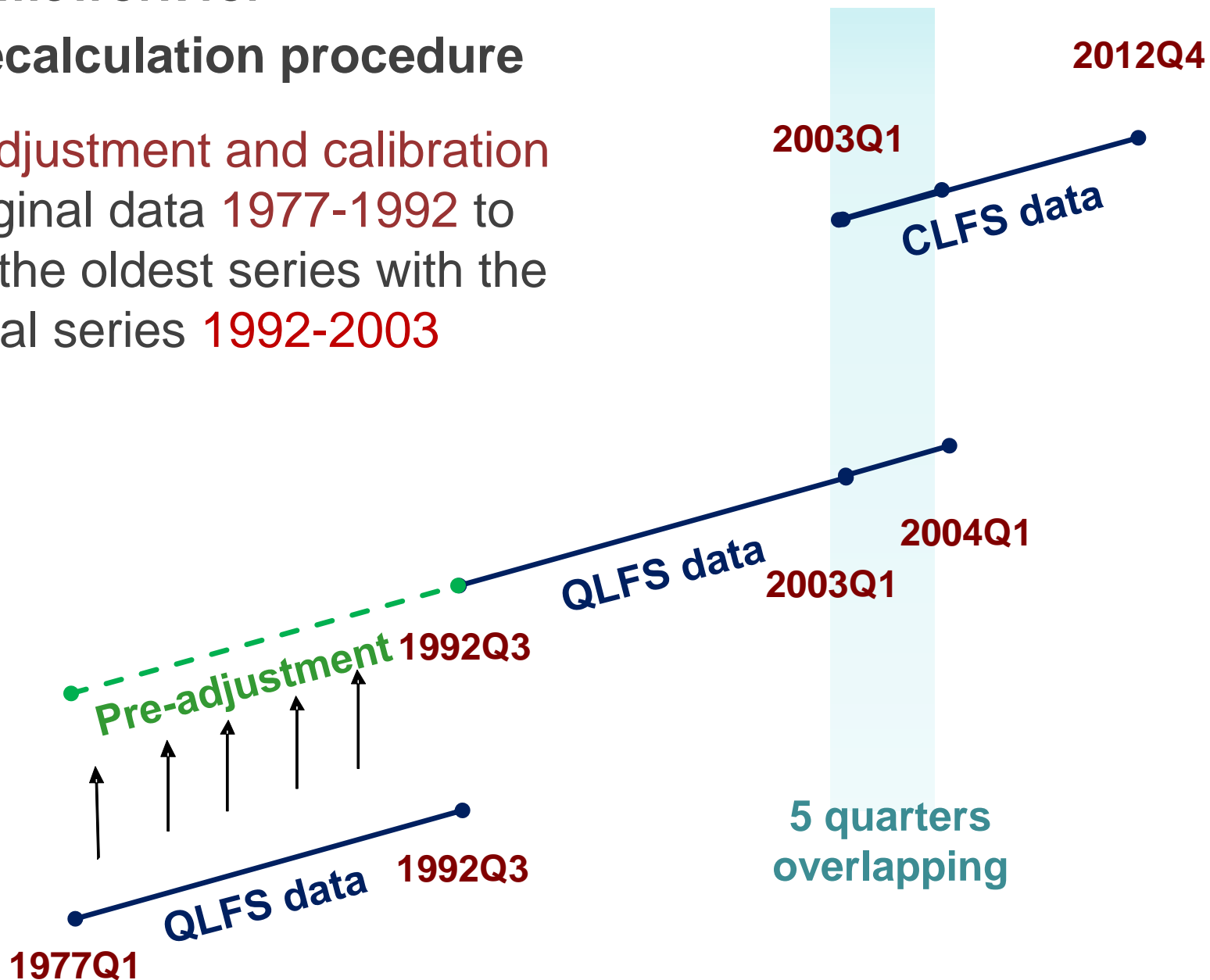**1992Q3**
2003Q1
2004Q1

**QLFS data**
1977Q1
1992Q3

5 quarters overlapping

The framework for back recalculation procedure

⇨ pre-adjustment and calibration of original data 1977-1992 to align the oldest series with the original series 1992-2003

2012Q4

2003Q1

CLFS data

2004Q1

QLFS data   **2003Q1**

Pre-adjustment   **1992Q3**

QLFS data   **1992Q3**

1977Q1

5 quarters overlapping
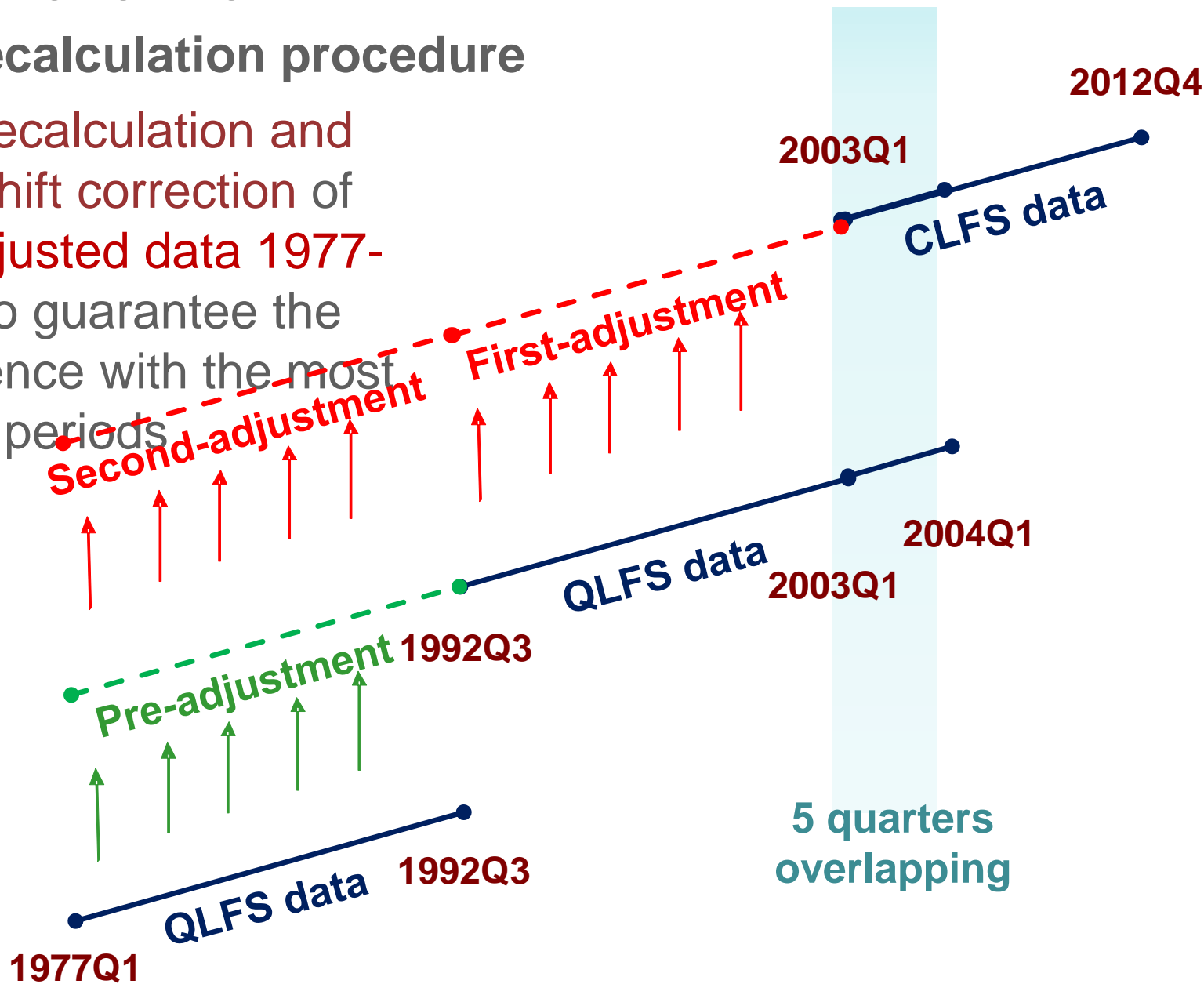
The framework for back recalculation procedure

⇨ Back recalculation of LFS series was already available for the period 1992-2003

2012Q4

2003Q1

CLFS data

First-adjustment

QLFS data   2003Q1

2004Q1

Pre-adjustment 1992Q3

QLFS data   1992Q3
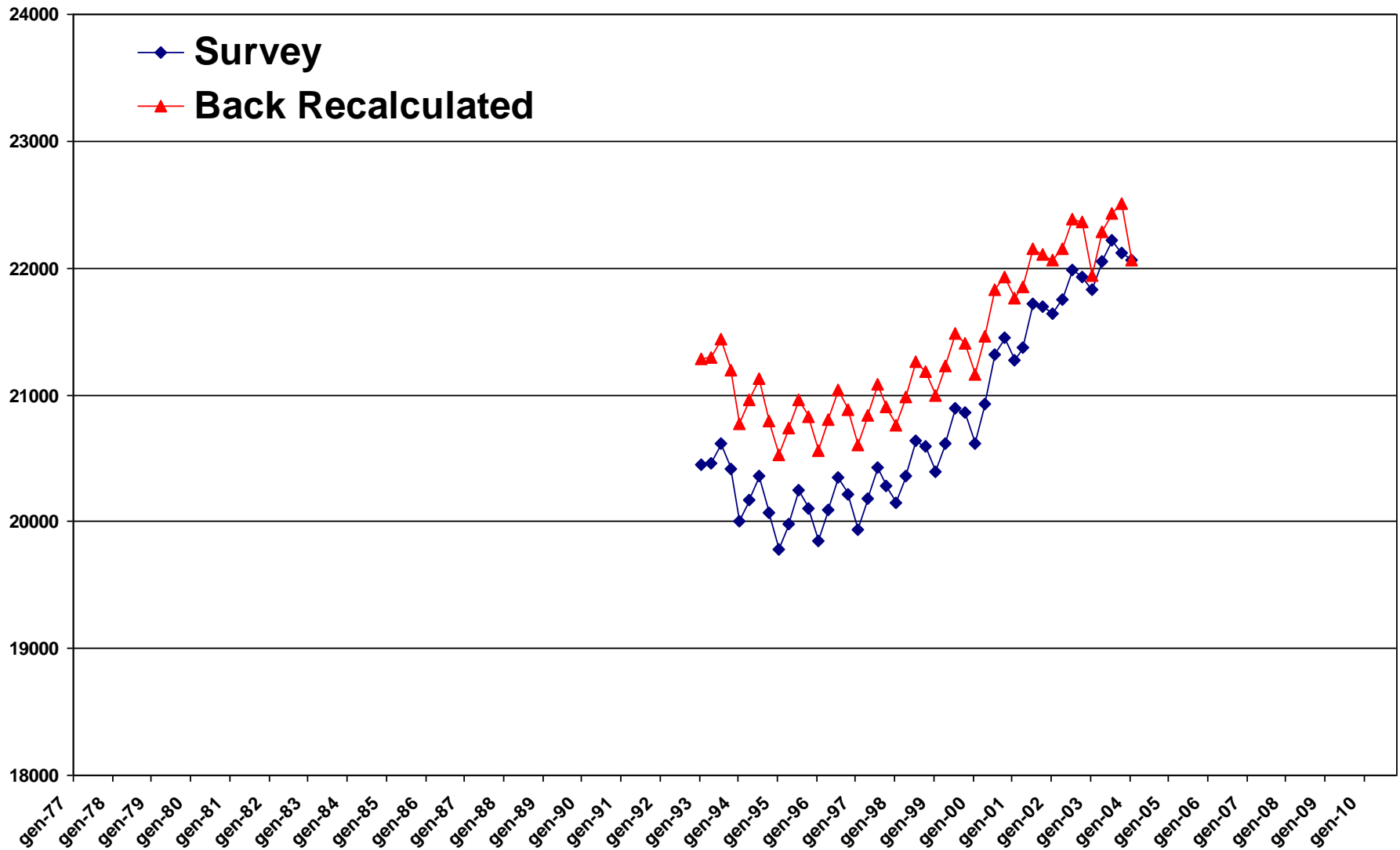
1977Q1

5 quarters overlapping
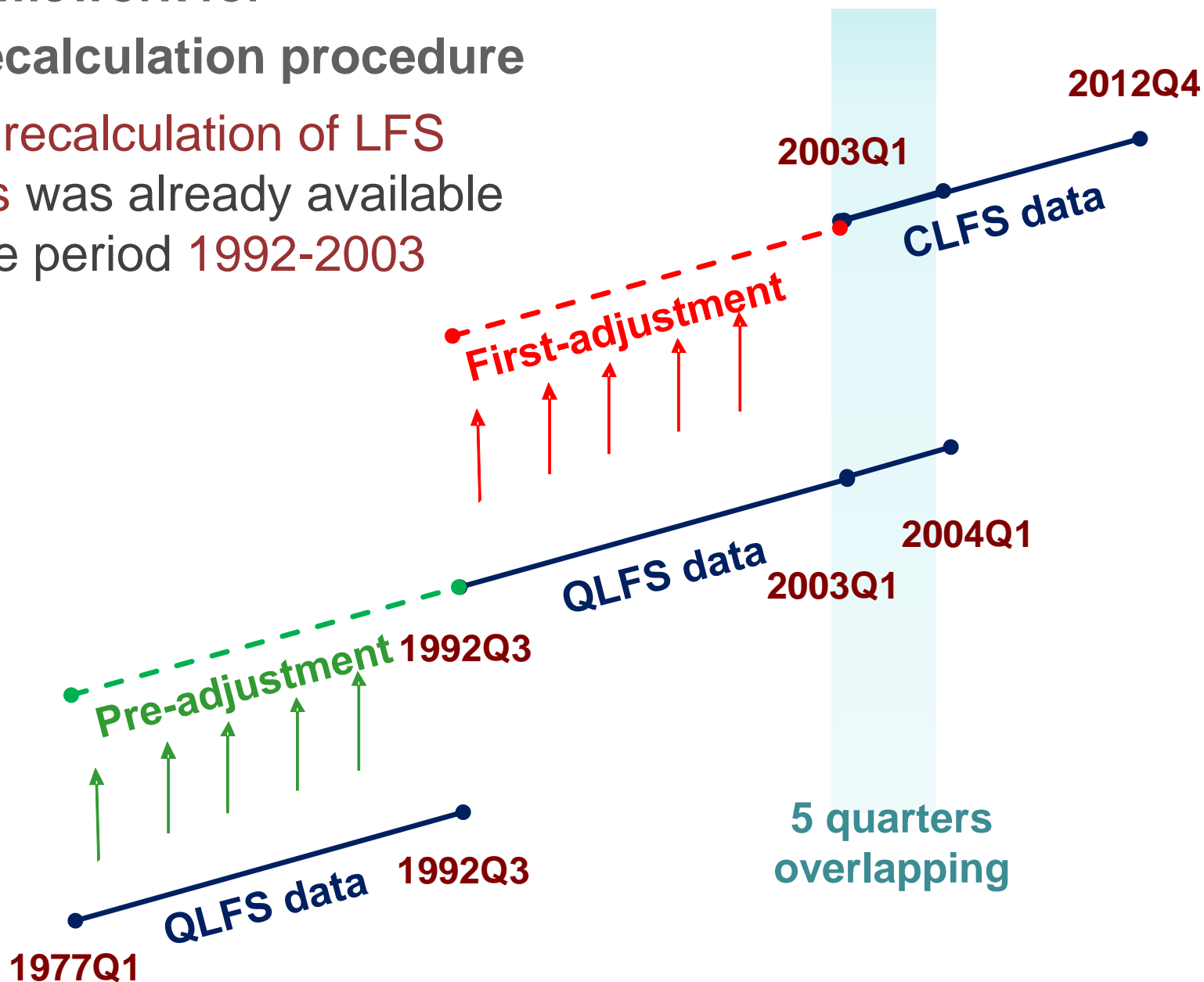
The framework for back recalculation procedure

⇨ back-recalculation and level-shift correction of pre-adjusted data 1977-1992 to guarantee the coherence with the most recent periods

2012Q4

2003Q1

CLFS data

First-adjustment

Second-adjustment

QLFS data

2004Q1

2003Q1

Pre-adjustment 1992Q3

QLFS data

1992Q3
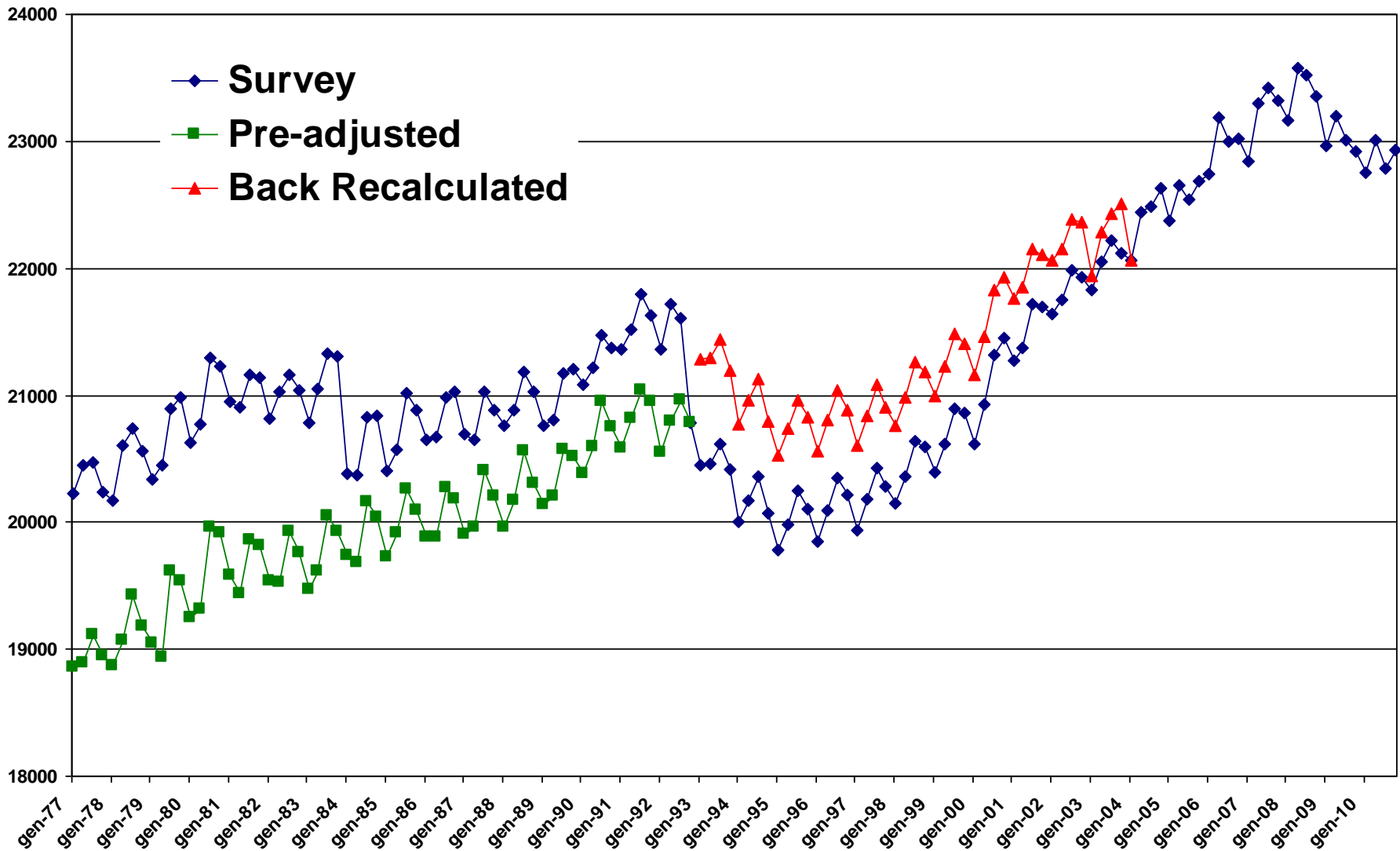
5 quarters overlapping

1977Q1

Employment 1992 - 2004
(absolute values in thousands)

Employment 1977 - 2010
(absolute values in thousands)

Employment 1977 - 2010
(absolute values in thousands)

Legend:
- Survey
- Pre-adjusted
- Back Recalculated

X-axis labels: gen-77, gen-78, gen-79, gen-80, gen-81, gen-82, gen-83, gen-84, gen-85, gen-86, gen-87, gen-88, gen-89, gen-90, gen-91, gen-92, gen-93, gen-94, gen-95, gen-96, gen-97, gen-98, gen-99, gen-00, gen-01, gen-02, gen-03, gen-04, gen-05, gen-06, gen-07, gen-08, gen-09, gen-10

Y-axis labels: 18000, 19000, 20000, 21000, 22000, 23000, 24000
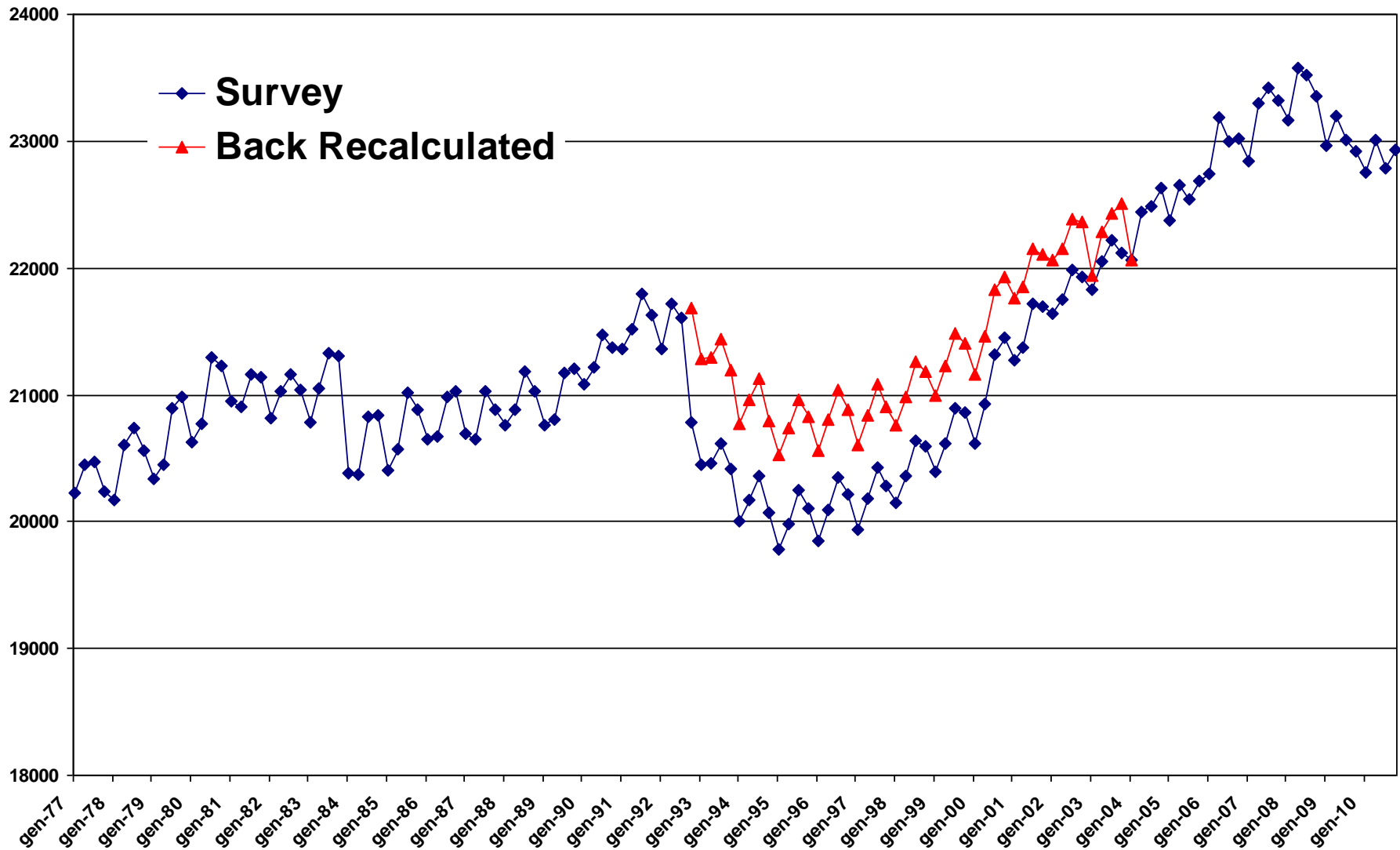
Employment 1977 - 2010
(absolute values in thousands)
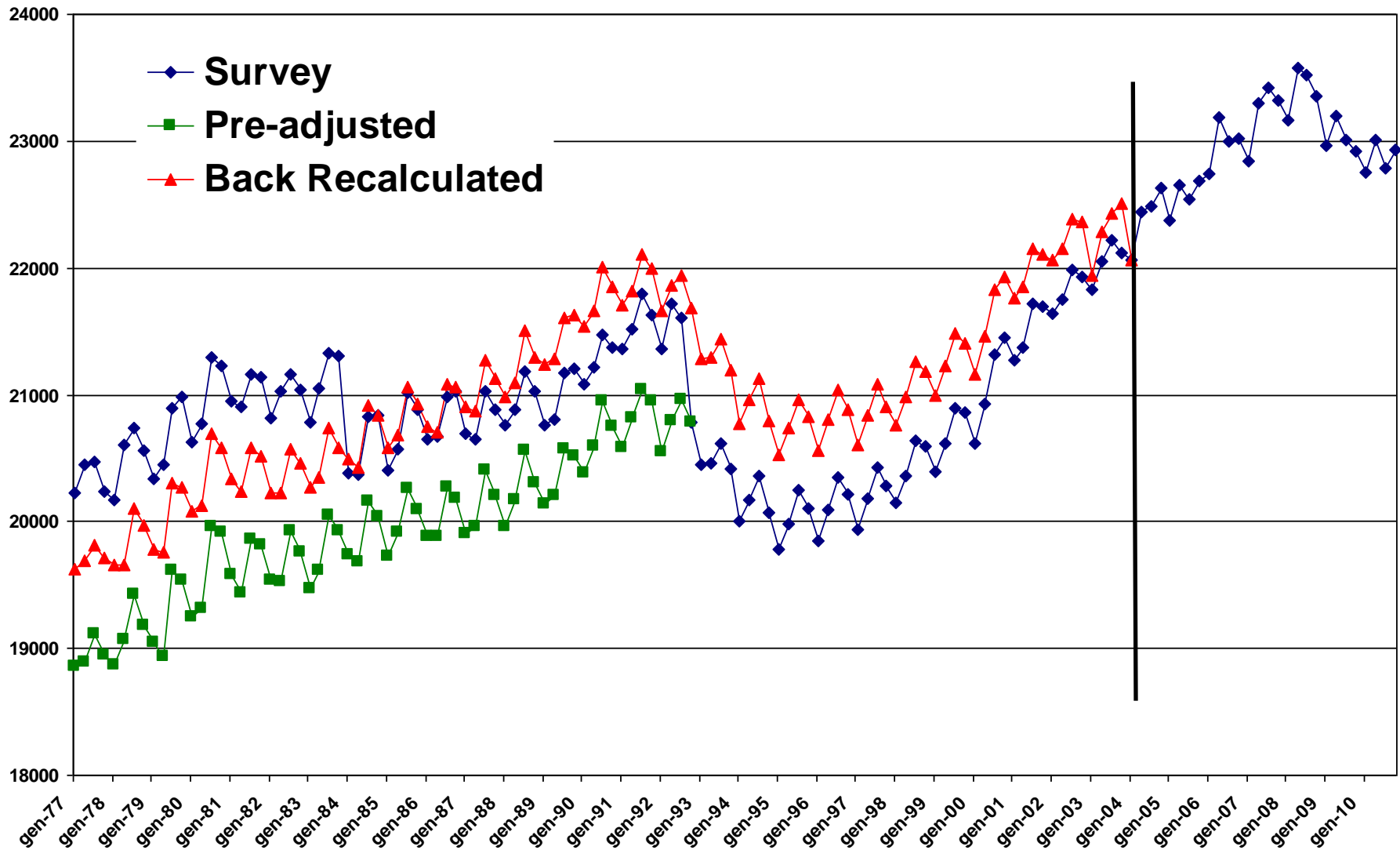
The transition to the new NACE Rev2 classification in 2008
(EC Regulation No 1893/2006)

The transition to the new NACE Rev2 classification in 2008 (EC Regulation No 1893/2006)

The project for transition to the new NACE started in 2007 to solve the following problem

From one side EC Regulation No 1893/2006
- ↳ established that NACE Rev.2 was to be used from 2008Q1 onwards

From the other side the Italian National Accounts System
- ↳ still required that employment is measured in NACE Rev1.1 up to 2010 (other 3 years) to guarantee the continuity of their estimates
- ↳ In 2011 will require quarterly time series of employment measured in NACE Rev2 from 2004 to back recalculate their aggregates.

Istat

Three years of overlap, between the two NACEs, to adjust the break of the LFS series

Thus, it was decided to simultaneously code the variables related to economic activity, according to both NACE Rev. 2 and NACE Rev. 1.1 for three years, 2008-2010.

However, LFS interviewers were using the Italian version of the Classification of economic activities NACE Rev1.1, called Ateco2002, which consists of about 10.200 names of economic activities, available as "on line dictionary".

How to automatically get the NACE codes according to both Rev. 1.1 and Rev. 2 classifications?

During the overlap period, the national edition of NACE Rev. 1.1 was used as usual, but further details were added to the "online dictionary" in order to establish, for any name of economic activity, the complete correspondence with NACE Rev. 2 at 2 digits level.

Three years of overlap, between the two NACEs, to adjust the break of the LFS series

Recoding the names was not straightforward, as different types of correspondences between NACE Rev. 1.1 and NACE Rev. 2 can be distinguished:

1-to-1,   n-to-1,   1-to-m,   n-to-m.

About 82% of the names in Ateco 2002 have a correspondence 1-to-1 or n-to-1 with new NACE. In this case the new code was automatically assigned.

The remaining 18% (about 1.800 names), had multiple correspondences and it was necessary to revise them manually. A solution was found:

- either adding new descriptions for the new economic activities of NACE Rev2,
- splitting a single description of NACE Rev1.1 into two or more descriptions associated with different codes of NACE Rev. 2.

Istat

The «online dictionary» as a tool for NACE codes conversion

1-to-1 and n-to-1

NACE 1.1	NACE Rev.2
16.00 - Manufacture of tobacco products	12 - Manufacture of tobacco products
61.10 - Sea and coastal water transport	50 - Water transport
61.20 - Inland water transport	50 - Water transport

1-to-m and n-to-m

NACE 1.1	NACE Rev.2
92.20 - Production of radio and television programmes	59 - Motion picture, video and television programme production, sound recording and music publishing activities
92.20 - Broadcasting of radio and television programmes	60 - Programming and broadcasting activities

Such tool allowed us to obtain automatically and simultaneously the pair of codes by choosing the correct description of economic activity during interviews (interviewers didn't know this new codes).

Istat

Differences on employment in «Industry» and «Services» in NACE Rev1.1 and NACE Rev2 for per period 2008-2010
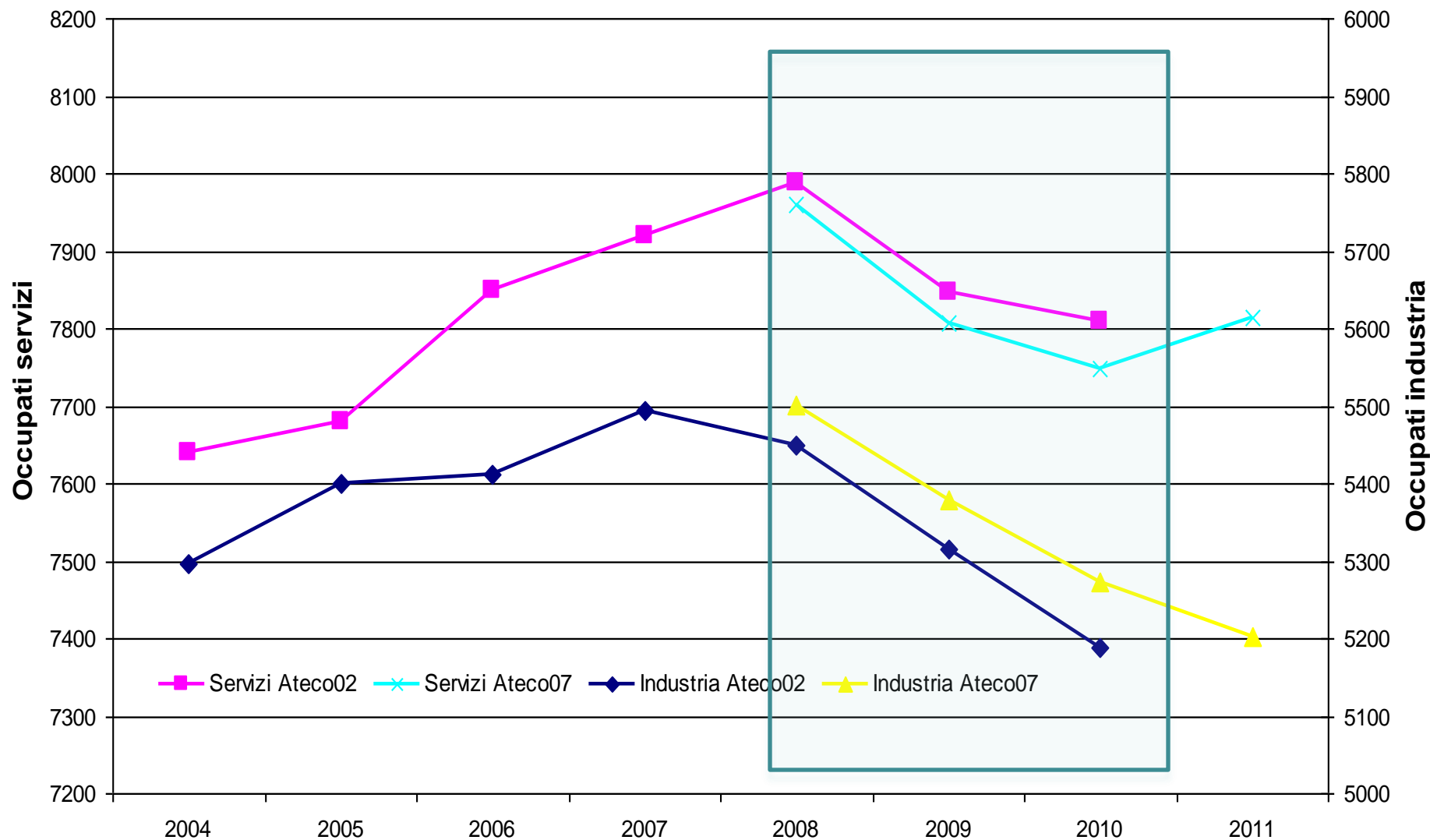
The method used for back recalculation

↳ Use the relationship captured during 2008-2010

↳ The problems of transition from the old to the new classification affected only the categories that have multiple correspondences.

It was used a techniques:

↳ based on the joint distribution

↳ allow to compute

- for each category of Ateco 2002 (old national classification) ,
- that disaggregates into "n" categories of Ateco 2007 (new national classification),
- the "weights" of those n categories (incidence)

↳ Apply these weights to each category of Ateco 2002 for previous period

Istat

Results

The application of this method resulted in:

⇨ a comprehensive set of employment data for the period 2004-2010, containing estimates in both Ateco 2002 and Ateco 2007, by gender, NUTS2, professional status and job duration (used by National Accounts).

⇨ a restricted set of historical series for the period 1977-2010 to be used for seasonal adjustments of current series.

Istat

The transition to new ISCO08 in 2011

As for the new NACE classification:

- Eurostat requires that LFS data on occupations have to be collected according to the new Isco08 from 2011.

- Hence, with the first quarter of 2011, Istat adopted the new national classification of occupations (called CP2011), which is completely linked to Isco08.

- With the adoption of the new Cp2011 and the change in the taxonomy ISTAT introduced a new approach to occupational coding.

Istat

The new approach to occupational coding

In fact, LFS moved from

↳ an occupational titles-oriented logic

to

↳ an approach which uses a detailed description of the activities actually performed by the worker,

The "on-line dictionary" used by interviewers in the electronic questionnaire to code occupations have consequently changed:

↳ the trigram search within the alphabetical list of occupational titles

has been replaced

↳ by an hierarchical search engine which focuses on the work activities actually performed by the respondent.

Istat

On line dictionary for ISCO08 coding

Inserisci la professione che v   Cerca

3.3.1.3.2 - Intervistatori e rilevatori professionali

LA POSIZIONE NELLA CLASSIFICAZIONE

3 - PROFESSIONI TECNICHE

3.3 - Professioni tecniche nell'organizzazione, amministrazione e nelle attività finanziarie e commerciali

3.3.1 - Tecnici dell'organizzazione e dell'amministrazione delle attività produttive

3.3.1.3 - Tecnici del trasferimento e del trattamento delle informazioni

3.3.1.3.2 - Intervistatori e rilevatori professionali

ESEMPI DI PROFESSIONI

o intervistatore professionale

o rilevatore professionale

3.3.1.3.2 - Intervistatori e rilevatori professionali

Le professioni comprese in questa unità assistono gli specialisti nella ricerca e nella acquisizione di informazioni, ovvero conducono interviste strutturate e semi strutturate con questionari e strumentazioni complesse in indagini e rilevazioni totali o campionarie disegnate su basi scientifiche.

ESEMPI DI UNITÀ PROFESSIONALI AFFINI CLASSIFICATE ALTROVE

o 3.3.1.3.1 - Tecnici dell'acquisizione delle informazioni
o 4.3.2.4.0 - Addetti ai servizi statistici

Similar occupations

Istat - Istituto Nazionale di Statistica
Via Cesare Balbo 16 00184
Roma tel. +39 06 46731
Progetto e sviluppo: Alessandro Capezzuoli

examples

Detailed description of the work activities

5 digits hieranchical classifications

Back recalculation of Occupational data

The change of the coding strategy and of the classification made time-series recalculation quite awkward.

It was not possible to use the same strategy of NACE

An ex-post recoding was made for all occupations of the interviews collected in 2011, according to the 3-digit of the previous national classification (Cp2001).

The double coding was quite straightforward for the 70% of the occupational units, given the 1-to-1 relationship.

For the remaining 30%, the situation was much more blurred, especially when the classification effect's go beyond the information stemming from the occupational titles (see for instance the case of managers / supervisors).

Istat

To check the quality of Back recalculation

Longitudinal information about all the persons who did not change jobs (occupation) between 2010 and 2011 were instead used to check consistency between

↳ occupational ISCO88 code given in 2010

and

↳ the ISCO88 code recalculated in 2011

Correct matches is 89.7% (3 digits) 92.7% (1 digit)

The double coding for the year 2011 gives us the opportunity to apply the same methodology used for the economic activity to back re-calculate time series of occupations

Istat

THANK YOU FOR YOUR ATTENTION!

….. I'm joking !

Introduction of 3 new NUTS3 and modification of two NUTS2 boundaries in 2010

still to be presented.

Introduction of 3 new NUTS3 and modification of two NUTS2 boundaries in 2010.

The LFS disseminates a number of estimates broken down by the main socio-demographic characteristics, at NUTS 3 level as annual averages.

NUTS3 code is also available on micro-data file for research

Starting from January 2010,

↳ 3 new NUT3 "province" were created, *bringing their number from 107 to 110.*

↳ 7 municipalities moved from one NUTS2 region to another *(Marche to Emilia Romagna), and thus* from one NUTS1 to another *(Centro to Nord Est).*

Istat

How to solve the Problem ?

↳ National users wants immediately that LFS estimates reflect this change, from 1st quarter 2010

↳ Eurostat would not accept the new boundaries (new codes) until 1st quarter 2012,

What we did during those 2 years?

↳ Estimates were disseminated with the old boundaries up to 2011Q4, and with the new boundaries from 2012Q1.

but

↳ to give the possibility to conduct short term analysis, relating to the new territorial unit since from the moment of their creation, LFS Unit recalculated the estimates produced by the LFS for the 8 quarters of 2010 and 2001, using the new boundaries

The approach for back-recalculation

↳ **recoding of micro-data** (for all variables related to territorial classification)

↳ **recalculation of the grossing weights** taking into account **new known totals of population.**

The method used is based on the need to **simultaneously achieve** the following **two objectives:**

1. for territorial units (NUTS3, NUTS2 and NUTS1) not affected by changes of boundaries, the quarterly and annual estimates already disseminated should not change at all;

2. original national estimates need not change because the borders of the nation do not change and do not change the reference population as a whole.

Istat

The approach for back-recalculation

⇨ the new estimates were obtained by recalculating the weighting factors only for the part of the sample falling in the areas affected by the territorial changes. This was make separately for three distinct spatial domains

  - example: one of this domain refers to the old province of Milano that was split into "Milano" and "Monza".

⇨ recalculation of the grossing weights after putting an appropriate set of constraints in the calibration estimator, in addition to the usual quarterly ones, so that

a) the new NUTS3 estimates must reflect the new known total of resident population;

  The constraints necessary to achieve this objective are the same as the normal procedure for quarterly estimates (population of the domain by sex and 14 age groups; population of NUTS3 by sex and 5 age groups; non-nationals by sex and citizenship; Number of households of the domain by rotation group; etc).

Istat

Constraints in calibration procedure

b) new quarterly estimates, referring to each of the three domains as a whole, should remain the same as those already disseminated

Example: the sum of the estimates of the new provinces of Milan and Monza may not differ from the estimate of the old province of Milan

The additional constraints, that actually allow the achievement of the objective stated in b) *refer to:*

- all the indicators disseminated by the LFS, quarterly and annual, at NUTS1, NUTS2 and NUTS3 level, resulting cross-classifying the most relevant labour market and demographic variables (labour status, gender, educational level, age class, nationality, employment status, professional status, full-time part-time, sectors of economic activity, occupation, unemployment duration (short, long), previous experience, Neet.

↬ Overall, the calibration process incorporated about 2.500 simultaneous constraints.

Results of recoding and re-weighting

↳ When this recalculation was completed, in March 2013, and at the same time of the 2012Q4 press realease, LFS was able to:

  ↳ update all LFS indicators on datawarehouse I.stat,

  ↳ update all the tables of annual average on the website,

  ↳ provide to the users the new 8 quarterly micro-data files,

  ↳ provide new micro-data files to Eurostat in accordance with the obligations arising from Community changes in the NUTS classification.

  Since then, users were allowed to analyze NUTS3 figures, referring to the new NUTS classification, starting from the year 2010.

Istat

THANK YOU FOR YOUR ATTENTION!

and VERY MUCH INDEED

for your

PATIENCE,

TOLERANCE,

TENACITY,

mental alertness,

physical resistance,

great capacity to remain calm ....although..

Istat