



7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

G. Data quality: Quality issues on monthly figures, timeliness and consistency between quarterly and monthly data

G1 – Quality improvement of monthly data in Romanian LFS

Lavinia Elena Balteanu – Romania

INS - Romania

Quality improvement of the LFS monthly data in Romania

Summary

Lavinia Elena Balteanu and Ruxandra Moldoveanu

7th Workshop on Labour Force Survey Methodology

Background

As a response to the high interest expressed by the users and employment policy makers, Romania has started few years ago to develop a methodology on unemployment figures on a monthly basis.

The model to estimate the ILO unemployment is based on a direct derivation of the estimates of quarterly LFS, without using any other auxiliary variables from other data sources.

The estimated monthly time series allow the assessment of short term unemployment trends.

Sampling plan

LFS sampling design has been founded on a two-stage sampling technique.

In the first stage, a stratified random sample of 780 areas, Primary Sampling Units (PSU) was design after the 2002 census, using as stratification criteria the residence area and county (NUTS 3 level) ¹.

In the second stage, 9360 clusters, composed of three housing unit each, are systematically selected from the initial sample of PSUs. The final sample consists of 28080 dwelling units per quarter. All households within each sampling unit are included.

The sample is built based on the dwellings rotation proceeding (rotational scheme 2-2-2), having as basic principle the following technique: a dwelling is surveyed for two successive quarters, it is temporarily taken out from the survey in the next two quarters, it is introduced again in the survey in the next two quarters, then it is taken out for good from the survey. Therefore, a dwelling is administrated in total for 6 quarters.

Field –work

- PAPI interviews;
- At territorial area level:
 - Collecting data from questionnaire and data entry
 - Applying checking rules and correcting errors
 - Data validation
 - monthly databases are complete and transferred from regional offices to the central level for final processing at M+28
- At INS – headquarter level:
 - Sampling
 - Monthly data processing
 - Data validation
 - Data weighting
 - Data tabulation
 - Data dissemination
 - For monthly unemployment estimates
 - Receiving monthly micro-data
 - Data validation
 - Monthly weighting and data analysis
 - Applying the algorithm for monthly estimates
 - Seasonal adjustment
 - Quality assessment of monthly data
 - Data dissemination

¹ The Multifunctional Sample of Territorial area called the master sample EMZOT

Methodological approaches related to unemployment monthly data

First attempt to produce monthly estimates from the quarterly survey was done in 2006-2007 within PHARE Stat 2004 project - with financial support from Eurostat. Data series considered: 2002-2006.

Two possible approaches were tested at that time:

- genuine monthly estimates:
 - o weighting of the monthly sample using a procedure similar to that one used for the quarterly results but with reduced calibration scheme: NUTS 2 (8 regions) x residence area (urban/rural) x gender x age (15-24 years; 25 years and over);
- 3 monthly moving averages: creation of moving "quarters" which were weighted similar to regular quarter: NUTS 2 (8 regions) x residence area (urban/rural) x gender x 5 years age groups.

Quality issues with the estimates produced:

- genuine monthly estimates:
 - o small sample size with results of high volatility (especially for young unemployed);
 - o Data could be available only approximately at $M + 45$.
- 3 monthly moving averages:
 - o smoother series and consistent with quarterly path
 - o they are not real "monthly data" and "month" to "month" change must be interpreted with caution;
 - o If such an estimate is labeled as the middle month, than data would be available at $M + 75$.

Two problems to be solved:

- Timeliness ($M + 25$) - most important. Due to PAPI data collection, data are not available weekly but monthly → no data for month M are available before $M + 25$ (not even a partial sample) → data for month M have to be predicted based on the most recent information on trend development.
- Reduce volatility.

A new method to overcome the drawbacks for the genuine monthly estimates

Prerequisite: a first improvement with no additional costs was done in 2008 → reallocation of the sample (within the strata) on the reference weeks so that the monthly samples become more balanced at NUTS 2 level.

Algorithm of the methodology developed is as follows:

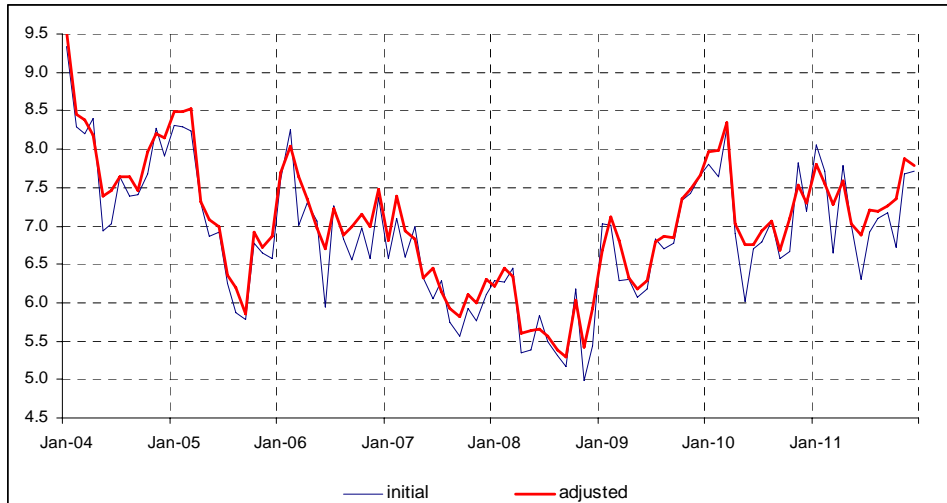
- For the estimated time series on ILO monthly unemployment, the data weighting procedure is similar with that used for quarterly LFS data; the only differences are:
 - o The procedures are applied for a sub-sample (monthly instead of quarterly): the initial sample is divided into 3 sub-samples corresponding to the 3 months of the quarter
 - o monthly samples are weighted by applying a reduced calibration scheme: gender X 4 age groups X area of residence X regions (NUTS2 level) X total number of households → initial monthly estimates.
- Based on the initial estimation (for each month), an improved (adapted to the purpose) version of **Holt method** is applied; this exponentially smoothes data series that show a linear trend → monthly time series of projected estimates

- In order to obtain an idea about the results of the applied method, the estimates are compared to a benchmarked data series (i.e. the original ones coming from the regular LFS) – monthly forecast time series (not seasonally adjusted)
- Adjustments (SA and trend cycle) are performed by using DEMETRA – TRAMO/SEATS.

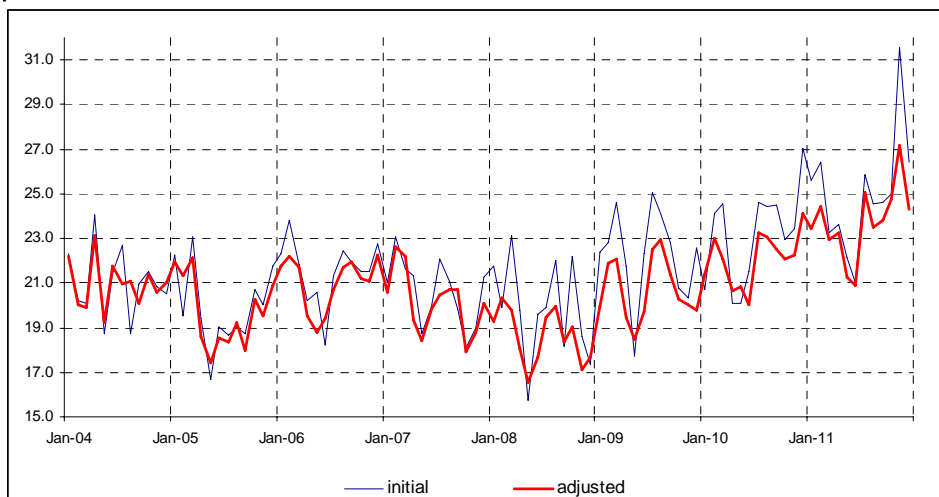
The unemployment rates are calculated from the adjusted series on unemployed persons and active population.

Results - examples

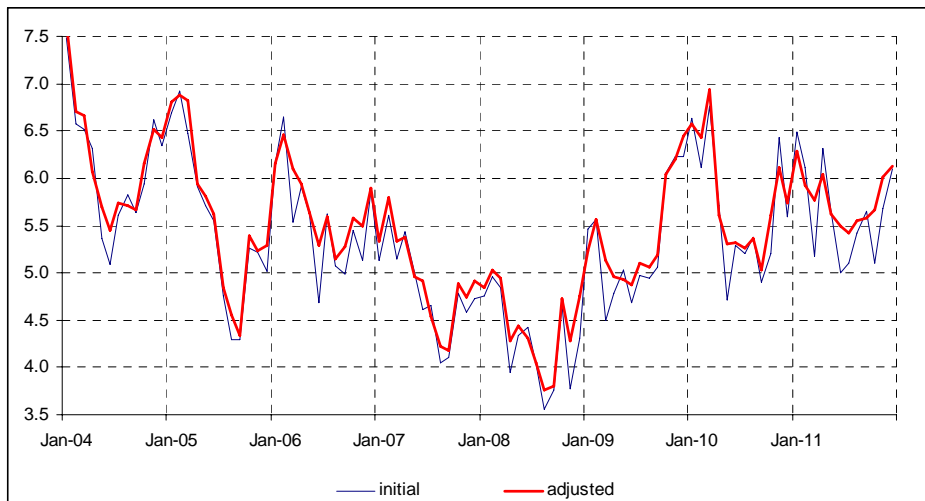
Unemployment rate - total



Unemployment rate - young persons (15-24 years) – not reliable (not published)



Unemployment rate - adults (25-74 years)



Revision policy

- Quarterly, after finalizing quarterly result from LFS, by benchmarking (e.g.: in May the estimates for January, February and March are revised and April is released as forecasted value).
- Annually, re-estimation of the model used for seasonal adjustment → the revision the entire series (seasonally adjusted and trend) are revised.

NSA

		Unemployment – thou pers		Unemployment rate %	
		First release	Revised	First release	Revised
2011	June	697207	681059	7.1	6.9
	July	690794	717590	7.0	7.2
	August	679734	713606	6.9	7.2
	September	684317	723664	6.9	7.3
	October	708214	718886	7.2	7.4
	November	712947	770405	7.2	7.9
	December	697442	763964	7.1	7.8

SA

		Unemployment – thou pers		Unemployment rate - %	
		First release	Revised	First release	Revised
2011	June	727276	725073	7.5	7.4
	July	707101	731736	7.3	7.5
	August	708864	728770	7.3	7.5
	September	727308	748014	7.5	7.7
	October	726944	721906	7.3	7.3
	November	726863	750215	7.3	7.6
	December	703224	751422	7.0	7.5

		T			
		Unemployment pers	– thou	Unemployment rate -%	
		First release	Revised	First release	Revised
2011	June	728374	73277 0	7.5	7.5
	July	713498	73211 9	7.3	7.5
	August	715023	73585 0	7.4	7.5
	September	722535	73697 7	7.5	7.5
	October	735291	73500 0	7.5	7.5
	November	718386	73701 3	7.3	7.5
	December	703821	73573 4	7.1	7.4

Quality assessment

- Analyzing of the estimates from one month to the other and from a quarter to the other as well as comparing with the same period of the previous year
 - o By gender
 - o By age groups (less than 25 years, 25-74 years)
- Analyzing of the monthly estimates by comparing with the quarterly results
- Analyzing of the LFS micro-data for ensuring the data accuracy and availability at the requested deadline
- Ensuring the comparability with definitions and concepts used at EU level.

Dissemination

- Main indicators obtained are:
 - ILO unemployed – total (15-74 years) by gender and age groups (15-24 years and 25-74 years), and
 - ILO unemployment rate – total (15-74 years) by gender and age groups (15-24 years and 25-74 years).
 - 3 series disseminated for each indicator:
 - non-seasonally adjusted,
 - seasonally adjusted,
 - trend
- Focus – on SA series (in press release); NSA and T series – only in the online database.*
- Because of the small number of cases of observation, the reliability of estimates for the indicators corresponding to the category of young people (age group 15-24 years) is extremely low; the series obtained showing a high degree of volatility. Therefore, the indicators corresponding to the category "youth" are not disseminated.

- Data disseminated by:

- Sending to Eurostat at M + 25 days.

First transmission in July 2011 (with reference period - June).

- Press release – monthly at M + 30 days

The first release of data was performed on 1st of August 2011 - press release - with monthly data series for:

- *January 2004 - March 2011 - as benchmarked values;*
- *April, May, June 2011 – as predicted values.*

- TEMPO on-line database
- Monthly Statistical Bulletin

Conclusions:

The possible econometric models that can be used provide information on the behavior of the phenomenon, but we should be aware of the limits due to the forecast model.

In the model used for Romania:

- Existence of the forecast errors – differences between estimates and real values of the raw series
- Deviations between forecasted values and values of the reference series (benchmark).

Even not so accurate than the quarterly data, the monthly estimates provide more rapid information that is requested by the users.