Professor Vergil VOINEAGU, PhD
E-mail: Vvoinagu@insse.ro
President of National Institute of Statistics, Romania
Silvia PISICA, PhD
National Institute of Statistics, Romania
Nicoleta CARAGEA, PhD
National Institute of Statistics, Romania

FORECASTING MONTHLY UNEMPLOYMENT BY ECONOMETRIC SMOOTHING TECHNIQUES

Abstract. *In the last year, the need for high-quality statistics for labour policy became increasingly significant. Requirements of European official statistics for short-term indicators on labour market statistics created the opportunity to develop a national methodology for deriving monthly unemployment estimates directly from the Labour Force Survey (LFS) results. LFS is carried out by Romanian National Institute of Statistics on quarterly basis, according to the European regulations and methodology. The unemployment is measured in line with International Labour Organization criteria, ensuring full comparability at European level. In response to this requirement Romanian National Institute of Statistics has developed such a methodology that generates the forecast of monthly unemployment rate using data from the series of Labour Force Surveys.*

The journey for reaching the present model was a complex and difficult one. The main part of the study is aiming to present comprehensively the work performed for developing the actual method used for regularly producing the monthly unemployment estimates. The outputs of the method are presented in the second part of the paper showing the usefulness brought by the model. It also emphasizes the seasonality of the time series suggesting an improvement of the Holt method. Main unmet expectations of the methodology are closing the present paper.

Key words**:** unemployment, labour force survey, estimation, forecasting

JEL Classification: C32, C53, J21

1. Introduction

A new approach to forecasting the trend and seasonality in time series data is presented. It is based on the Labour Force Survey results using the relationship between the historical and forecasted data. Concerning this approach, three forecasting data series were produced in order to smooth the values of data series and also to exhibit fairly good accuracy. The preliminary results show that for the trend and seasonal data, the proposed methods produce better forecasts than the other methods tested.

2. Literature Review

A time series is a sequence of observations which are ordered chronologically. In the collection of data taken over time it is inherent to capture some forms of random variation. Special methods for reducing the random variation effects are used, of which "smoothing" techniques most widely. These methods, when properly applied, reflect more clearly the underlying trends. At the end of 50's, Holt and Winters proposed a method for smoothing and forecasting time series with locally linear trend (Winters, 1960) nowadays referred usually as Holt method. They employed the exponentially weighted moving average approach to estimate the level and newly also the slope of the time series. Several smoothing techniques were also developed and described by Jelinek and Mercer (1980), Katz (1987), Bell, Cleary, and Witten (1990), Ney, Essen, and Kneser (1994), and Kneser and Ney (1995).
The main methods of time series smoothing are presented in Box 1.

Box 1. The main smoothing techniques

The simplest smoothing technique is to take an average of past values. Unfortunately, this also completely obscures any trends, changes, or cycles within the data. More complicated averages eliminate some but not all of this obscuring and still tend to lag as forecasters, not responding to changes in trends until several observations after the trend has changed.

Moving Averages: Moving averages rank among the most popular techniques for the pre-processing of time series. They are used to filter random "white noise" from the data, to make the time series smoother or even to emphasize certain informational components contained in the time series. In other words, the moving average technique only uses the most recent observations or a weighted average that values some observations more than others. Exponential smoothing represents an attempt to improve upon these defects.

Exponential Smoothing: This is a very popular scheme to produce a smoothed Time Series. Whereas in Moving Averages the past observations are weighted equally, Exponential Smoothing assigns exponentially decreasing weights as the observation get older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

The benefit of simple exponential smoothing is that it allows for a trend in how the smoothed data is changing. It does poorly, however, at separating changes in the trend from the random variations inherent to the data. For that reason, double and triple exponential smoothing are also used, introducing additional constants and more complicated recursions in order to account for trend and cyclical change in the data.

Double exponential smoothing - also known as **Holt's Linear Exponential Smoothing -** is a refinement of the popular simple exponential smoothing model but adds another component which takes into account any trend in the data. Simple exponential smoothing models work best with data where there are no trend or seasonality components to the data. When the data exhibits either an increasing or decreasing trend over time, simple exponential smoothing forecasts tend to lag behind observations.

Holt method has achieved a broad popularity among forecasters due to its simplicity and good performance. Several modifications of this original method have appeared in literature and are used in practice (T. Hanzak, 2008).

3. Description of the econometric model to forecast the monthly unemployment

The starting point of the monthly unemployment smoothing and forecast in the Romania's case is the Holt method, using two different parameters and allowing to apply the forecasting for series with trend. In classical model, the standard Holt's method is given by:

$$\mathrm{Y_t} = \alpha y_t + (1-\alpha)(Y_{t-1} + \mathrm{b}_{t-1})$$

$$b_t = \beta(Y_t - \mathrm{Y}_{t-1}) + (1-\beta)b_{t-1}$$

where:

$\mathrm{Y_t}$ - is the smoothed level of the series, computed after y_t is observed;

$\mathrm{b_t}$ - is the smoothed trend at the end of period *t;*

$\mathrm{y_t} = \mathrm{Y_t} + b_t$

$\alpha > 0$

$\beta < 1$ - are the smoothing parameters.

Holt's two parameters exponential smoothing method is an extension of simple exponential smoothing. It adds a growth factor (or trend factor) to the smoothing equation as a way of adjusting for the trend.

The initialization process for Holt's linear exponential smoothing requires two estimates:

- One to get the second smoothed value for the level of the series;
- The other to get the trend b_2.

Start values used to initialize the model are the followings:

$\mathrm{Y}_2 = \mathrm{y}_2$

$\mathrm{b}_2 = \mathrm{y}_2 - \mathrm{y}_1$.

Forecast **m** periods into the future are given by:

$$F_{t+m} = Y_t + mb_t$$

where:
m - represents the periods to be forecast into the future.
Using an improved Holt method, it was produced the first estimator for the unemployment provided by Labour Force Survey, considered the reference data.
The reference time series are obtained dividing the initial quarterly sample of LFS in three sub-samples, one for each month of the quarter and each monthly sub-sample is grossed-up separately, applying a reduced calibration scheme.
The estimator is given by:

$$Y_t = \alpha y_t + (1-\alpha)(Y_{t-2} + b_{t-2})$$
$$b_t = \beta (Y_t - Y_{t-2}) + (1-\beta) b_{t-2}$$

where:

Y_t - is the smoothed level of the series, computed after y_t is observed;

b_t - is the smoothed trend at the end of period *t;*

$\mathrm{y}_t = \mathrm{Y}_t + b_t$

$\alpha > 0$
$\beta < 1$ - are the smoothing parameters.

In this case, the initialization process for Holt's linear exponential smoothing requires two estimates:

- One to get the first smoothed value for the level of the series;
- The other to get the trend b_1.

Start values used to initialize the model are the followings:

$\mathrm{Y}_1 = \mathrm{y}_1$ and $\mathrm{b}_1 = 0$.

Also assuming: $\mathrm{Y}_2 = \mathrm{y}_2$.

The first equation corrects the level of time series Y_t with the last two trend values b_{t-2}. The improvement of the classical method helps to determine the predicted values for the end of the quarter on the basis of the real values registered at begins of the quarter.
Smoothing coefficients $\alpha > 0, \beta < 1$ are determined by minimizing a measure of forecast error such as MSE (Minimum Square Errors):

$$\frac{1}{T}\sum_{t=0}^{T-1}\left(Y_{t+1} - \hat{Y}_{t+1}\right)^2 = \frac{1}{T}\sum_{t=0}^{T-1} e_{t+1}^2 \rightarrow \min$$

The forecast error is given by:

$$e_{t+1} = Y_{t+1} - \hat{Y}_{t+1}.$$

In order to asses the quality of the estimations of the proposed method, the results were compared with the benchmarked[1] data series, given by equation:

$$Y_t^{hmb} = \hat{Y}_t \times Y_Q^b / \overline{\hat{Y}_Q}$$

where:

Y_t^{hmb} - represents the benchmarked data series;

Y_Q^b - represents the quarterly average of the reference data series (LFS quarterly values provided from the sample);

$\overline{\hat{Y}_Q}$ - represents the quarterly average of the Holt's estimation series.

The prediction is given by:

$$Y'_t = \alpha\hat{Y}_{t-2} + (1-\alpha)Y_{t-1}^{hmb}$$

Forecast h periods into the future are given by:

$$Y'_{t+h} = \alpha\hat{Y}_{t+h-2} + (1-\alpha)Y_{t+h-1}^{hmb}$$

where:

h - are the periods to be forecast into the future.

On the base of this algorithm, the forecast for the first next period (h=1) is calculated as following:

$$Y'_{t+1} = \alpha\hat{Y}_{t-1} + (1-\alpha)Y_t^{hmb}.$$

The results obtained consist in monthly series of final estimates in non-seasonal adjusted form.

Beside the monthly series (number of unemployed persons and unemployment rate) adjusted for level and trend (non-seasonally adjusted series), seasonally adjusted (thus, removing the effect of seasonal variations) and trend (which represent the series from which, both, the seasonal and irregular effects, were removed) series are also calculated.

Seasonal adjustment was performed using DEMETRA software (TRAMO/SEARTS method), which also estimates the calendar effect (Orthodox Easter, leap year and other national holidays) and identifies and corrects the outliers (occasional transitory or permanent changes in level).

The estimation of unobserved components: trend-cycle, seasonal and irregular component was made with SEATS program based on ARIMA models. Seasonally adjusted series were obtained by removing the seasonal component from the original data. Trend was obtained by removing the irregular component from the seasonally adjusted series.

[1] Benchmarked data series are the original estimates obtained regularly (i.e. quarterly from the entire sample) from the Romanian LFS

4. Application of the model

The model was applied on the labour market data provided by LFS for the period 2004-2011. To gross-up the data for each month, the following variables are used for calibration: population by sex and four age groups (less than 15 years, 15-24, 25-64, 65 and over), area of residence (urban/rural), regions (NUTS 2 level) and total number of individual households.

Beside the monthly series (number of unemployed persons and unemployment rate) adjusted for level and trend (non-seasonally adjusted series), seasonally adjusted (thus, removing the effect of seasonal variations) and trend (which represent the series from which, both, the seasonal and irregular effects, were removed) series are also calculated. Thus, monthly series are built and disseminated in three forms: non-seasonally adjusted, seasonally adjusted, and trend. An exemplification of these data series is shown in the figures 1a and 1b.

Figure 1. a. Unemployment rate (15-74 years old)

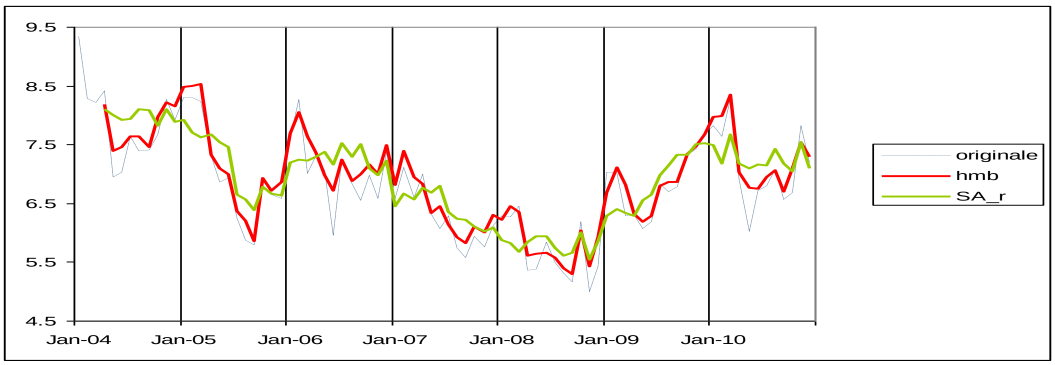
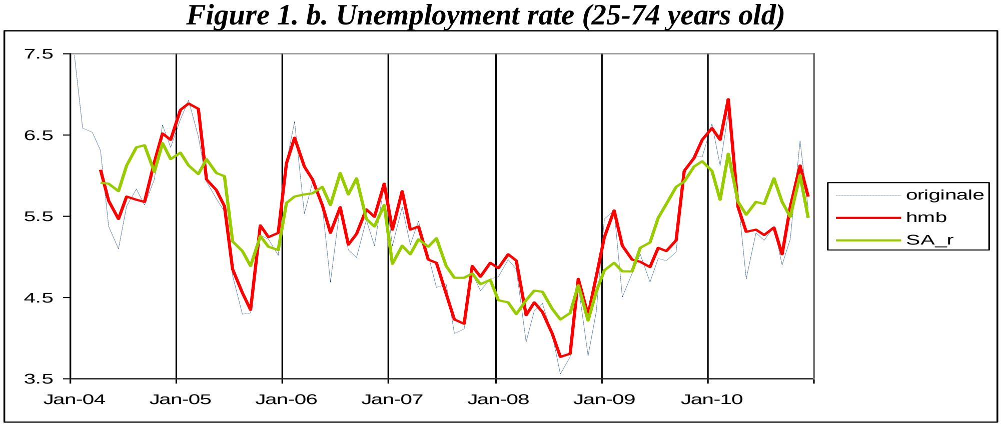


Figure 1. b. Unemployment rate (25-74 years old)



Source: *Authors' calculation based on LFS data, National Institute of Statistics, 2012.*

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 monthly estimates for "youth" category are not disseminated. The first release of data was performed on 1st of August 2011.

Negative impact of recession on employment

The unemployment rate is an important indicator with both social and economic dimensions. From an economic perspective, unemployment could show unused available labour. Increasing unemployment may be also reflected in loss of income for individuals and rising pressure on government spending on social benefits.

The unemployment rate according to the International Labor Organization (ILO) definition is the most widely used labor market indicator because of its international comparability.

ILO unemployed are the persons aged 15-74 years who simultaneously meet the following three conditions: (i) have no job, (ii) are available to start working in next two weeks, (iii) were actively seeking a job at any time during the last four weeks. ILO unemployment rate is the proportion of ILO unemployed in the total economically active population.

The official statistics in Romania calculate and disseminate three unemployment rates: (i) registered unemployment – with monthly periodicity, at NUTS3 (county) level, defined according to national legislation, provided by administrative sources – National Agency for Employment and Labour Force; (ii) ILO unemployment (according to International Labour Office criteria) – with quarterly periodicity (and as annual average), at NUTS 2 (region) level, defined according to International Labour Office (ILO) criteria, provided by statistical sources – Labour Force Survey (LFS), National Institute of Statistics (INS); (iii) monthly ILO unemployment - with monthly periodicity, at NUTS 0 (country) level, defined according to ILO criteria, provided by statistical sources – LFS and estimation model. In order to provide input for short-term employment policy, the European Community Statistical Office (EUROSTAT) requested Member States to develop their own models to estimate the ILO unemployment by direct derivation of the results of quarterly Labour Force Survey (LFS), without using any other auxiliary variables obtained from other data sources.

In the next section the monthly unemployment estimates are analyzed for period January 2004-January 2012, emphasising the impact of the economic crisis on the labour market.

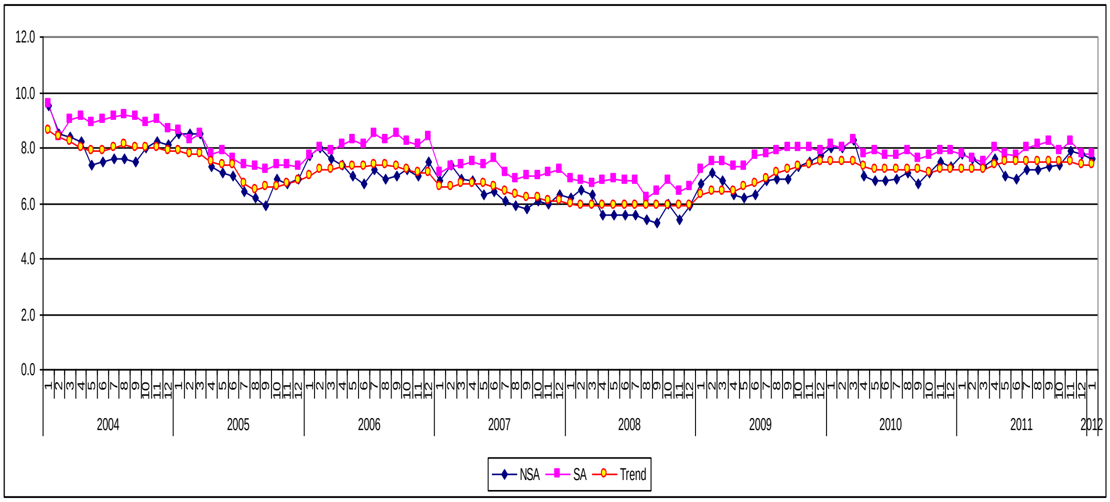
The Romanian economy knew a continuous positive GDP growth during 2000-2008, the unemployment significantly decreasing, when the labour market followed one of the economical principals: the unemployment tends to fall when growth is high. In that circumstance, the unemployment rate (calculated for 15-74 age group) decreased from 9.6 percents in the first month of 2004 to 6.2 percents in August 2008, when the value was the lowest during the entire analysed period. The middle of the year 2008 is the starting point of the global economic and financial crisis that affected also most of the countries across the world. Romania was no exception. As expected, mid of 2008 was an inflexion moment marking the installation of the unemployment increase. The unemployment started to grow especially until the first quarter of 2010 but not as much as in other European countries also affected by the crisis. Several reasons can explain the phenomenon.

On one hand, Romania has a quite strong share of underground economy, and consequently a high share of employment in the informal economy. As studies (Vasile, Pisică, 2011) shown one of the effects of the economic crisis was a shift of employment from formal into informal area, the share of persons engaged in informal sector growing in three years by 6 percentage points (from 23.2 percents in 2008 to 29.2 percents in 2010. On the other hand, Romania is an emigration country. It is noticed that, a factor influencing the employment dimension in Romania is the international migration, especially in the group of working age population (15-64 years), the phenomenon increasing since 2007 after joining the European Union. As the other Member States of EU declared to Eurostat[2] the flows of Romanian citizens[3] arriving in those countries (and of course leaving Romania) grew from less than 200 thousand persons in 2005-2006 to more than half million in 2007 and over 300 thousand persons in 2008 (after that phenomenon slowing down quite significantly since, because of the financial crisis, the job opportunities became very limited in the old Member States).

Monthly unemployment figures are based on results of the model presented in the section before (figures come from the seasonally adjusted series – SA, with unemployment rates calculated for both age groups 15-74 and 25-74 years old figure 2.a and 2.b).

[2] Eurostat – Statistical Office of the EU
[3] Eurostat database, http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home

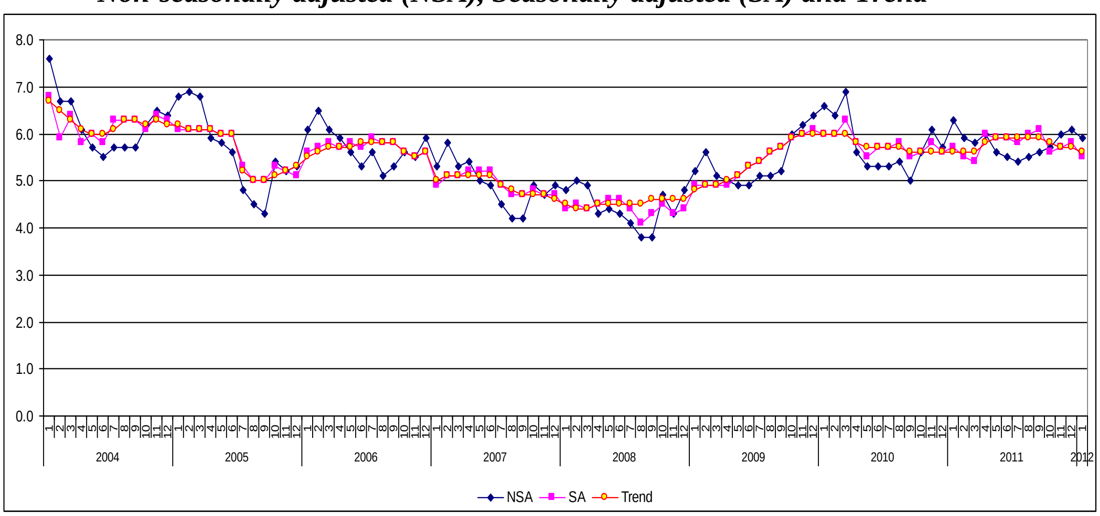
Figure 2a. Monthly unemployment rate (15-74 years) - % -
Non-seasonally adjusted (NSA), Seasonally adjusted (SA) and Trend



Source: *Authors' calculation based on LFS data, National Institute of Statistics, 2012*

The unemployment rate has been rising since August 2008, reaching the highest value in March 2010 (8.3 percents – for 15-74 age group, respectively 6.3 percents – for 25-74 age group).

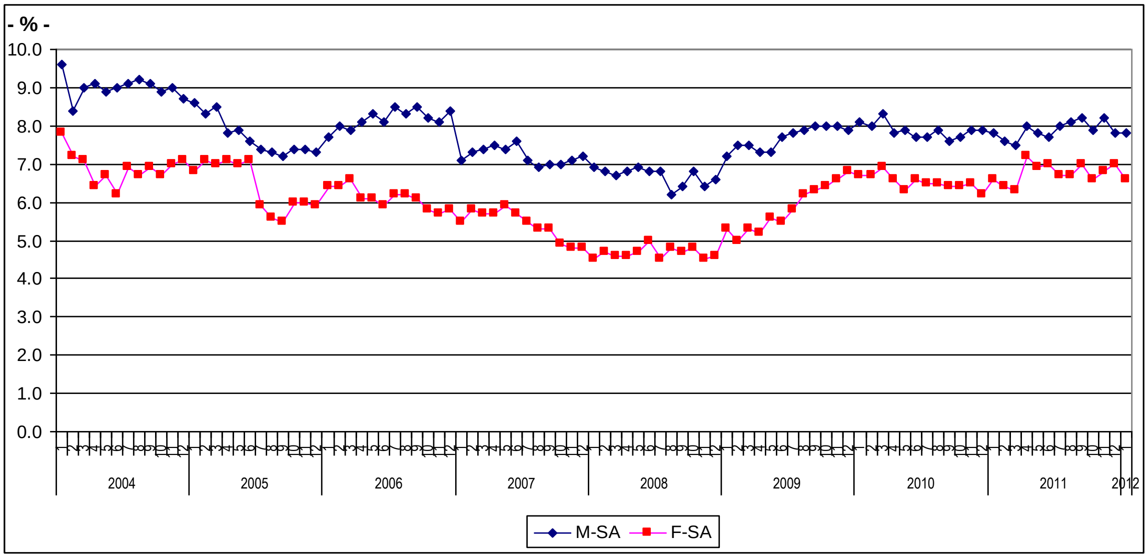
Figure 2.b. Monthly unemployment rate (25-74 years) - % -
Non-seasonally adjusted (NSA), Seasonally adjusted (SA) and Trend



Source: *Authors' calculation based on LFS data, National Institute of Statistics, 2012*

Most part of the unemployment belongs to the male population. The share of men in total unemployed persons varies from 56.7 percents in June 2005 to almost two thirds, i.e. 66.1 percents in January 2008.

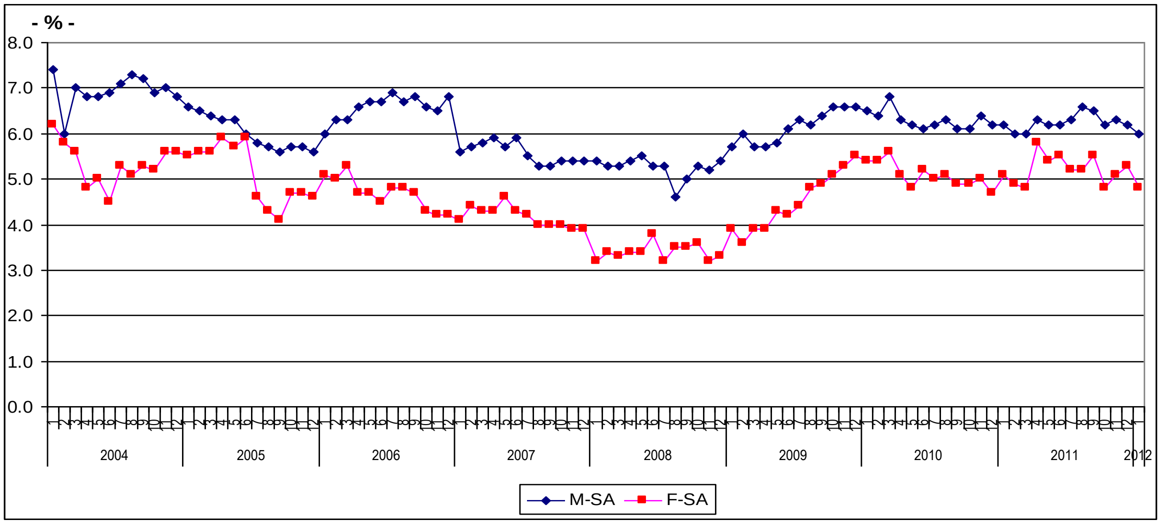
Figure 3a. Monthly unemployment rate by gender (15-74 years) - % - Seasonally adjusted (SA)



Source: *Authors' calculation based on LFS data, National Institute of Statistics, 2012*.

One particularity of the Romanian labour market is the sense of the unemployment disparities by gender. In most Member States of the European Union unemployment is deeper among women as compared to men. In Romania is the opposite: the unemployment rate for male exceeds the one for female. The gap between the two varies for those aged 15-74 from 0.5 percentage points (June 2005) to 2.8 percentage points (June 2004) and for the group 25-74 years old from 0.1 percentage points (June 2005) to 2.6 percentage points (December 2006).

Figure 3b. Monthly unemployment rate by gender (25-74 years) - % - Seasonally adjusted (SA)



Source: *Authors' calculation based on LFS data, National Institute of Statistics, 2012.*

For the last couple of years (starting with the last months of 2009) the gap between the two rates of unemployment has the tendency of decreasing while the share of male in the total number of unemployed persons is more steadily close to 60 percents (if before 2010 the value of the share could vary within an interval of 10 percentage points, during last two years the interval decreased to 4 percentage points)..

Conclusions

Based on the one improved Holt's model, the method described in this study is the most suitable model for forecasting monthly unemployment rate in Romania. Each model type has unique characteristic which fits to a particular data series. More forecasting techniques should be explored to ensure fitness to longer series of unemployment rate.

There are several advantages that can be obtained when using double exponential smoothing technique:

- exponential smoothing models could fit with computer system; in this case, a simple program such as Microsoft Excel can be used to generate new forecasts;
- the models of exponential smoothing react more quickly to changes in data patterns than the moving average;

The main problem could be encountered when using this method is the determination of the size of smoothing coefficients. The criterion is to choose α such that the MSE is

minimum. With the assistance of solver facility available in Microsoft Excel, the amount of work is saved when searching for the best parameter α value.
Econometric models provide formulations and assumptions on variables subject to forecast, reflecting the orientation of the behaviour of the studied phenomenon.
Limits due to the forecast model are the followings:

- existence of the forecast errors, defined as the difference between estimates and real values of the raw series;
- the model impose the minimization of these errors by the condition that mean of squares of forecast errors to be minimal;
- deviations between forecasted values and values of the reference series (benchmark) due to their dependence on the trend of the previous period, and factors that can act unpredictably over economic environment. These deviations are larger in the case of youth aged 15-24 unemployment data series, the differences being exacerbated by the volatility of the series for this segment of the population.

REFERENCES

[1]**Hanzak, T**. **(2008),** ***Improved Holt Method for Irregular Time Series***, WDS'08 Proceedings of Contributed Papers, Part I, 62–67;
[2]**Holt, C.C (2004)*, Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages*** - Journal of Economic and Social Measurement, ISSN 1875-8932 (Online) http://iospress.metapress.com/content/pyu1rk484x1kb1qq/ ;
[3]**Jula, D., Jula, M. (2010)),** ***Metode de prognoza***, Mustang Publishing, Bucharest;
[4]**Lynwood, A. Johnson, D.(1990),** ***Forecasting and Time Series Analysis***, McGraw-Hill,Inc, 2nd edition;
[5]**Vasile, V., Pisica, S.(2011), Trends of Employment in Informal Sector in Romania during Crisis; Economic and Social Effects** ; 7th edition of the International Conference "EUROPEAN INTEGRATION – NEW CHALLENGES", Faculty of Economic Sciences – Oradea University, Oradea, 2011, ISBN 978-606-10-0521-5, p.496-514;
[6]**Voineagu, V., Caragea, N., Pisica, S.(2011),** ***Methodological Proposal for Compiling the ILO Unemployment with Monthly Periodicity ;*** Romanian Statistical Review no. 7 , p. 46-60;
[7]Winters, P. R. (1960), ***Forecasting Sales by Exponentially Weighted Moving Averages*;** *Management Science* 6, pp. 324-342;
[8]European Commission Eurostat, Directorate F: Social Statistics and Information Society, Unit F. Labour market, working documents http://circa.europa.eu/Members/irc/dsis/ employ/library?l=/&vm=detailed&sb
[9]Eurostat database, http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home
[10]Unemployment statistics *(Doc.: Eurostat/D1/LAMAS/11/05)*

[11]Harmonised monthly unemployment rates (Doc.: *Eurostat/F2/LAMAS/ 01 /07)*
[12]Monthly unemployment (Doc.: *Eurostat/F2/LAMAS/ 12 /08*)
[13]Harmonised monthly unemployment rates (Doc.: *Eurostat/F2/LAMAS/ 07 /06)*
[14]Revised methodology for the calculation of monthly unemployment rates (*Eurostat/F2/LAMAS/ 32a /08)*
[15]Feasibility study on fl ash estimates of monthly unemployment rates using EU sampling on weekly LFS samples (*Doc.: Eurostat/F2/LAMAS/ 32b /08)*
[16]Institutul National de Statistică (INS) România, metadata database http://colectaredate.insse.ro/metadata/public.htm?locale=ro