Proceedings of ALGORITMY 2005
pp. 334–340

MODELLING SLOVAK UNEMPLOYMENT DATA BY A NONLINEAR LONG MEMORY MODEL *

JOZEF KOMORNíK † AND MAGDA KOMORNíKOVÁ ‡

Abstract. The main visual feature of Slovak monthly unemployment data (from the period January 1993 – August 2004) is a dramatic increase after the elections in October 1998 which can be expressed by a step in the deterministic level function. In the residuals from this level function we can identify a significant cyclical component and a long memory structure (with the estimated value of the Hurst parameter close to 0.8), as well as an autoregressive short memory behavior.

Key words. Time series modelling, smooth regime switching, spectral analysis, fractional integration, ARFIMA modelling.

AMS subject classifications. 62M10, 62M15, 62M20

1. Introduction. Analysis of US month unemployment data (from the period July 1968 – December 1999) based on a nonlinear long memory model has been performed in [Van Dijk *et al.*, 2000]. The considered nonlinearity (with 2 regimes) was based on the observation that US unemployment seems to grow faster in recessions than it falls in expansions.

The main visual feature of Slovak monthly unemployment data (from the period January 1993 – August 2004) is a dramatic increase after the elections in October 1998 (and following changes in composition of government coalition parties and in economic policies, that had supported artificial employment in unrentable businesses and economic expansion based on a dramatic increase of foreign indebtedness). This switch of political regime corresponds to a parallel switch of "regime" in our time series, which can be expressed by a remarkable positive shift in its values (the average value for the period before the end of 1998 was 13.41 %, for the following period 17.11 %). Visual investigation of the values of our time series around the end of 1998 indicates that the increase has been quick but not totally instant. Inspired by the methods of smooth transition regime-switching models, we model this change in values by a logistic class smooth step function that we will call *base level function*. In the residuals from this base level function, we can identify a significant cyclical component and a long memory structure (with the estimated value of the Hurst parameter close to 0.8), as well as an autoregressive short memory behavior.

2. Methods and results of the modelling Slovak Unemployment data. According to modern principles of time series modelling (see Franses, 1998) we adapt individual steps of our analysis to the main visual properties of the studied time series of monthly values of unemployment rates in Slovak Republic in the period January 1993 – August 2004 that is shown below.

2.1. Regime switching change of base level. Since the level of Slovak unemployment have clearly risen after the election (in end subsequent change of Govern-

* The research summarized in this paper was partly supported by the Grants GACR 402/04/1026 and VEGA 1/0273/03.

† Faculty of Management, Comenius University, Bratislava, Slovakia

‡ Faculty of Civil Engineering, Slovak University of Technology, Slovakia and UTIA AV CR Prague, Czech Republic

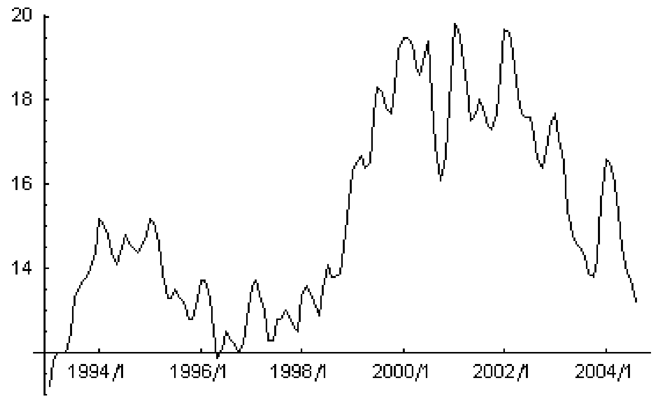

FIG. 2.1. *Slovak unemployment data from the period January 1993 – August 2004*

ment) in October 1998 (corresponding to $t = 70$), we tried to model the consequence of this change by a base level function

$$L(t) = a + bG(t, c, \gamma) \tag{2.1}$$

where

$$G(t, c, \gamma) = \frac{1}{1 + e^{-\gamma(t-c)}} \tag{2.2}$$

is a logistic function.

Applying methods of nonlinear optimization we received the following (least squares) parameter estimates:

$$a = 13.362; b = 3.738; c = 71.36; \gamma = 2.73.$$

The shape of the resulting base level function is shown at the following Fig. 2.2.

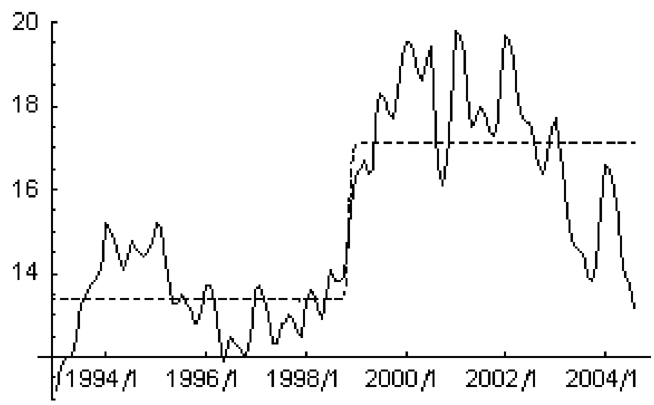

FIG. 2.2. *Original time series and the base level function.*
Original data —, Base level function - - -

2.2. Identification of seasonal and cyclical component. Now we proceed in analysis of residuals of original unemployment data from corresponding values of the estimated base level function. We return to classical procedures of time series analysis and estimate the coefficients (and test significance) of the seasonal (annual) period component in this residuals. The contribution of this seasonal component (with the amplitude close to 0.5) to the explanation of the variability of the analyzed time series is strongly significant ($p < 0.01$). Next we proceed by calculating the values of periodogram for remaining residuals (Tab. 2.1)

$$\hat{f}\left(\omega_j\right) = \frac{\hat{\gamma}\left(0\right)}{2\pi} + \sum_{k=1}^{N-1} \hat{\gamma}\left(k\right) \frac{\cos\left(k\,\omega_j\right)}{\pi}$$

where N is the length of the time series ($N = 140$), $\omega_j = \frac{2\pi j}{N}$ is j-th Fourier frequence and $\hat{\gamma}\left(k\right)$ is the sample covariance function for lag k.

Table 2.1
Periodogram for remaining residuals after base level function and seasonal component

N/j [months]	$\hat{f}\left(\omega_j\right)$	**N/j [months]**	$\hat{f}\left(\omega_j\right)$
70	8.83	140/13	0.52
47	2.85	140/12	0.48
140	1.96	140/9	0.44
35	0.93	140/24	0.42
6	0.57	140/8	0.36

Applying step-wise regression procedure on the relation between the considered residual time series and periodic functions with the above Fourier frequencies, we conclude that the most important periods are (approximately) 6, 4 and 3 years.

The resulting deterministic function (systematic component) has the form

$$\begin{aligned} X_t = {} & 13.36 + \frac{3.74}{1 + e^{-2.73(t-71.36)}} + 0.34\cos\left(\frac{2\pi t}{12}\right) + 0.36\sin\left(\frac{2\pi t}{12}\right) \\ & -0.53\cos\left(\frac{\pi t}{70}\right) - 0.73\cos\left(\frac{\pi t}{35}\right) - 0.51\cos\left(\frac{3\pi t}{70}\right) - 0.36\cos\left(\frac{2\pi t}{35}\right) \\ & -0.26\sin\left(\frac{\pi t}{70}\right) + 1.02\sin\left(\frac{\pi t}{35}\right) + 0.50\sin\left(\frac{3\pi t}{70}\right) + 0.18\sin\left(\frac{2\pi t}{35}\right) \end{aligned}$$

The quality of the fit of the original time series by the combined systematic component (base level function, seasonal and cyclical components) can be seen in the Fig. 2.3.

The graph of the autocorrelation function of residuals from the combined systematic component function (see Fig. 2.4) indicates the existence of significantly nonzero terms both of low and high orders.

2.3. Modelling long memory features. In order to identify and eliminate high order (long memory) terms, we try to find (for the considered time series of residuals Z_t) a representation of the ARFIMA form [Fouskitakis et al., 2000, Van Dijk et al., 2000]

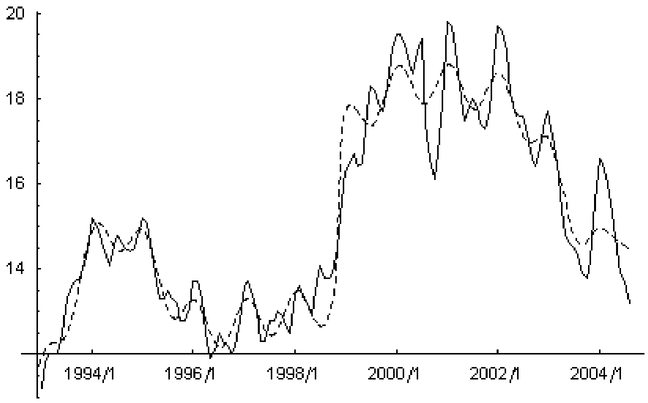

FIG. 2.3. *Original data and regression function. Original data —, Systematic component - - -*

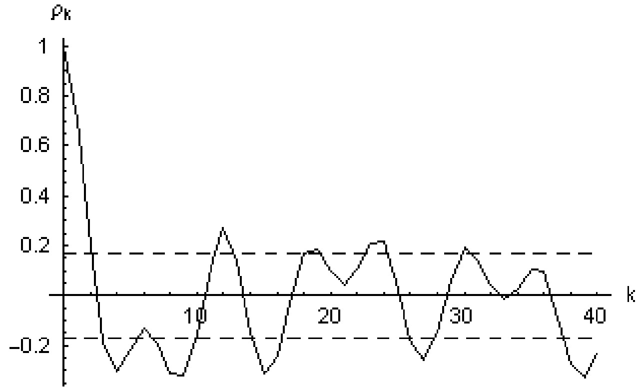

FIG. 2.4. *Autocorrelation function of residuals*

$$(1-B)^d Z_t = a_t \tag{2.3}$$

where

B is the backward shift operator (i.e. $BZ_t = Z_{t-1}$),

$\{a_t\}$ is a covariance stationary ARMA process.

The fractional difference operator is defined by

$$\begin{aligned}(1-B)^d &= \sum_{k=0}^{\infty}\binom{n}{k}(-B)^k \\ &= 1 - dB + \frac{d\,(d-1)}{2!}B^2 - \frac{d\,(d-1)\,(d-2)}{3!}B^3 + \cdots \end{aligned} \tag{2.4}$$

The series Z_t is covariance stationary if $d < 0.5$ and invertible if $d > -0.5$. The autocorrelation function of Z_t does not decline at an exponential rate, as is characteristic for covariance-stationary ARMA processes, but rather at a (much) slower hyperbolic rate. For $0 < d < 0.5$, Z_t possesses long memory in the sense that the autocorrelations r(k) are not absolutely summable. This implies that even though the $r(k)$'s are individually small for large lags k, their cumulative effect is important. The unknown parameter d can be estimated via so-called Hurst parameter $H = d + 0.5$ that is the slope in the regression

$$\log\left(\frac{R}{S}\right)_n = \log c + H \ \log n$$

for the so-called rescaled ranges $\left(\frac{R}{S}\right)_n$ (defined below). Details of the estimation procedures has been taken from [Millen and Beard, 2003].

Firstly the time series must be divided into D contiguous sub-series of length n, where $D \times n = N$, the total length of the time series. For each of theses sub-series m, where $m = 1, ..., D$:

1. Determine the mean E_m of each sub-series.
2. Determine the standard deviation S_m of each sub-series.
3. Normalise the data $\{Z_{i,m}\}$ by subtracting the mean from each data point:

$$X_{i,m} = Z_{i,m} - E_m, \quad \text{for } i = 1, \ldots, n.$$

4. Using the normalized data create a cumulative time series by consecutively summing the data points:

$$Y_{i,m} = \sum_{j=1}^{i} X_{j,m} \quad \text{for } i = 1, \ldots, n.$$

5. Using the new cumulative series find the range by subtracting the minimum value from the maximum value:

$$R_m = \max\{Y_{1,m}, \ldots, Y_{n,m}\} - \min\{Y_{1,m}, \ldots, Y_{n,m}\}.$$

6. Rescale the range, $\frac{R_m}{S_m}$ by dividing the range by the standard deviation.
7. Calculate the mean of the rescaled range for all sub-series of length n:

$$\left(\frac{R}{S}\right)_n = \frac{1}{D} \sum_{m=1}^{D} \frac{R_m}{S_m}.$$

8. The length of n must be increased to the nect higher value where $D \times n = N$ and D is an integer value. Step 1 to 7 are then repeated until $n = N/2$.
9. Finally, the value of H is obtained using an ordinary least squares regression with $\log(n)$ as independent variable and $\log\left(\frac{R}{S}\right)_n$ as the dependent variable. The slope of the resulting equation is the estimate of the Hurst exponent. The regression is run over values of n greater than 10, as small values of n produce unstable estimates when sample size are small.

The resulting value of the estimate of Hurst parameter (as the measure of long-term persistence of shocks) in our case is H = 0.792, which is significantly larger than the critical value 0.5. The estimate of the fractional differencing parameter d = 0.292, hence the series X_t is covariance stationary, invertible and possesses long memory.

We see that the quality of fit of the time series of residuals improved considerably after application long memory filter (see Fig. 2.5).

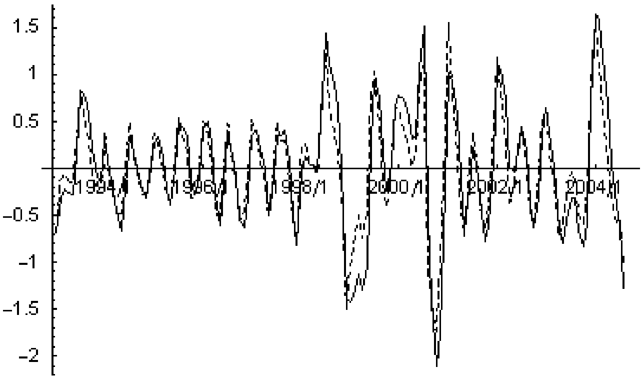

FIG. 2.5. *Time series of residuals and their model. Residuals —, model - - -*

2.4. Modelling short memory structure of residuals. As the last step of our analyzes we applied Box – Jenkins methodology to the residuals of the long term filter. The best fit (with respect to AIC and BIC criteria) has been received for a model in the class AR(7) (the numbers in parentheses are estimates of standard deviations of estimates of model coefficients):

$$Z_t = \underset{(0.08)}{0.92} Z_{t-1} - \underset{(0.11)}{0.46} Z_{t-2} - \underset{(0.09)}{0.14} Z_{t-3} - \underset{(0.09)}{0.15} Z_{t-5} + \underset{(0.11)}{0.31} Z_{t-6} - \underset{(0.08)}{0.44} Z_{t-7}$$

The final fit of original data is in the Fig. 2.6.

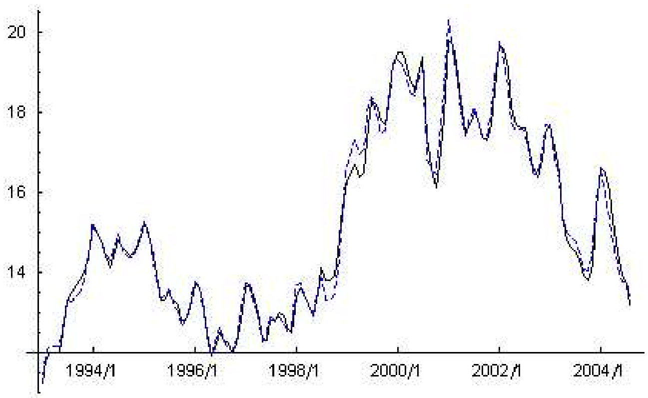

FIG. 2.6. *Time series of Slovak unemployment data and model*

The diagnostic check of the ultimate residuals (see Fig. 2.7 (left)) indicated no significant autocorrelations (see Fig. 2.7 (right)).

3. Conclusion. We selected four steps of analysis according to basic properties of the considered time series that have been manifested in the graphically visible change of the basic level after October 1998, in the periodogram of remaining residuals, in the pox diagram and in autocorrelation structure of resulting residuals. The resulting model exhibits a high order of fit and desired quality of diagnostic check for residuals.

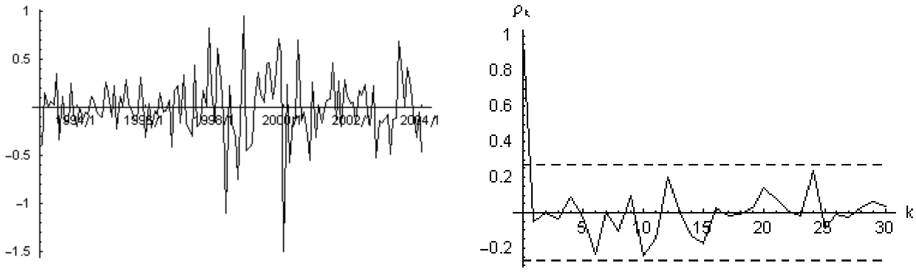

Fig. 2.7. *Residuals after model (left). Autocorrelation function (right).*

REFERENCES

[1] Artl, J., Artlová, M., *Finanční časové řady – Vlastnosti, metody modelování, příklady a aplikace.* GRADA Publishing, 2003.

[2] Fouskitakis, G. N., Fassois, S. D., *On the estimation of long-memory time series models.* University of Patras, Greece, 2000.

[3] Franses, P. H., *Time series models for business and economic forecasting.* Cambridge University Press, 1998.

[4] Franses, P. H., Dijk, D., *Non-linear time series models in empirical finance.* Cambridge University Press, 2000.

[5] Granger, C. W. J., Teräsvirta, T., *Modelling nonlinear economic relationships.* Oxford University Press, 1993.

[6] Millen, S., Beard, R., *Estimation of the Hurst exponent for the Burdekin River using the Hurst-Mandelbrot Rescaled Range Statistic.* Working paper on the University of Queensland, 2003.

[7] Rose, O., *Estimation of the Hurst Parameter of Long-Range Dependent Time Series.* Report No. 137, University of Würzburg, Germany, 1996.

[8] Van Dijk, D. , Franses, P. H., Paap, R., *A nonlinear long memory model for US unemployment.* Research Report EI2000-30/A, Econometric Institute, Erasmus University, Rotterdam, 2000.