The Second Order Logistic Smooth Transition Autoregressive Model for Unemployment Rate of Latvia

Oksana Pavlenko
Institute of Applied Mathematics
Riga Technical University
Riga, Latvia
Oksana.Pavlenko@rtu.lv

Liana Zeltina
Institute of Applied Mathematics
Riga Technical University
Riga, Latvia
lianaz3@inbox.lv

***Abstract*—To account for structural changes in data, the second order logistic smooth transition autoregressive model (LSTAR2) is obtained as the best smooth transition model using popular R software for the differences of the unemployment rate among 15-75 years old residents of Latvia. The specially adjusted R functions for estimation and forecasting of LSTAR2 based on the available in R package tsDyn functions for the first order logistic smooth transition autoregressive model (LSTAR1) are used. The constructed LSTAR2 model also is compared with the best chosen linear autoregressive, multiplicative seasonal autoregressive, self-exciting threshold and LSTAR1 models. LSTAR2 is superior than the compared models for these data, which indicates that the new R functions may be useful for economic data analysis.**

Keywords—time series, autoregressive model, threshold autoregression, smooth transition autoregressive model

I. Introduction

Different time series models are used to forecast macroeconomic and financial variables. Starting from trend, additive seasonal and linear autoregressive models, more and more advanced models are appearing. Different threshold and smooth transition models were added to linear autoregressive integrated moving average (ARIMA), multiplicative seasonal ARIMA models (SARIMA), Vector Autoregression (VAR) and Vector Error Correction (VEC) models. Over the recent years, we have seen that structural shifts in the economic situation are common. As a result, models with thresholds may describe the observed dynamics of economic variables better. Moreover, the transition among regimes is often gradual rather than instantaneous. Most popular models capturing this are delineated in the next section.

Smooth transition and threshold autoregressive model applications can be found in [1], [2], [3], [4] and many more.

II. Definitions and Estimation of Smooth Transition Models

A. *Definitions*

Let us define the models which will be used later.

Linear Autoregressive Integrated Moving Average model [5] ARIMA(p,d,q):

$\nabla^d Y_t = a_0 + \sum_{i=1}^{p} a_i \nabla^d Y_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t,$

where (here and further)

Y_t is our time series;

p, d, q are the orders of the model,

$\nabla^d = (1-\Delta)^d -$ difference operator.

Multiplicative Seasonal Autoregressive Integrated Moving Average model [6] SARIMA(p,d,q)(P,D,Q)$_s$:

$$\Phi_P(\Delta^s)\phi(\Delta)\nabla_s^D \nabla^d Y_t = \delta + \Theta_Q(\Delta^s)\theta(\Delta)\varepsilon_t,$$

where

s is the number of seasons,

P,D,Q are the orders of a seasonal part,

$\Delta Y_t = Y_{t-1}$ simple difference (lag),

$\phi(\Delta) -$ autoregressive operator,

$\theta(\Delta) -$ moving average operator,

$\nabla_s^D = (1-\Delta^s)^D -$ seasonal difference operator,

$\Phi_P(\Delta^s)$ =1- $\Phi_1\Delta^s$- $\Phi_2\Delta^{2s}$ -... - $\Phi_P\Delta^{Ps}$ – seasonal autoregressive operator,

$\Theta_Q(\Delta^s)$=1+ $\Theta_1\Delta^s$+$\Theta_2\Delta^{2s}$+ ... + $\Theta_Q\Delta^{Qs}$ – seasonal moving average operator.

Self-Exciting Threshold Autoregressive model [5] SETAR:

$$Y_t = \begin{cases} a_0^{(1)} + \sum_{i=1}^{p_1} a_i^{(1)} Y_{t-i} + \varepsilon_t^{(1)}, & if \quad Y_{t-d} \le r_1, \\ \dots & \\ a_0^{(k)} + \sum_{i=1}^{p_k} a_i^{(k)} Y_{t-i} + \varepsilon_t^{(k)}, & if \quad r_{k-1} < Y_{t-d} \le r_k \end{cases}$$

where

$p_1,\dots, p_k$ are orders of regression equations in corresponding regimes,

d is the delay parameter and

$r_1,\dots, r_k$ are thresholds.

Smooth Transition Autoregressive model [5] STAR:

$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_p\, Y_{t-p} + G(\gamma, x, th)\big[\beta_0 + \beta_1\, Y_{t-1} + \cdots + \beta_p\, Y_{t-q}\big] + \varepsilon_t$

where

$G(\gamma, x, th)$ – is a transition function,

γ is a smoothness coefficient,

th is a threshold value,

x is a threshold variable (some lag of the dependent variable Y like Y_{t-1} or may be their combination or some exogeneous variable.

Most popular transition functions are the following.

- <u>First order Logistic Smooth Transition Autoregressive model (LSTAR1)</u>

$G(\gamma, x, th) = \frac{1}{1+e^{-\gamma(x-th)}}$,

- Exponential Smooth Transition Autoregressive model (ESTAR)

$G(\gamma, x, th) = 1 - e^{-\gamma(x-th)^2}$, γ>0,

- Second order Logistic Smooth Transition Autoregressive model (LSTAR2)

$G(\gamma, x, th_1, th_2) = \frac{1}{1+e^{-\gamma(x-th_1)(x-th_2)}}$

where th_1 and th_2 are threshold values.

For more detailed description of the transition functions of STAR models, see [1].

B. *Estimation*

Smooth transition autoregressive models, as well as self-exciting threshold autoregressive models usually are estimated with nonlinear (conditional) least squares. Also, maximum likelihood estimation under assumption of normally distributed errors can be used for STAR models. Both methods are equivalent in this case [1,2].

Conditional on $(\gamma, th1, th2)$, the estimates of equations' coefficients α_i, β_i can be estimated by Least Squares. The parameters $(\gamma, th1, th2)$ are obtained using 3-dim grid search by minimizing the residual variance [2].

Different nonlinear optimization methods can be used to minimize the sum of squared residuals. The Newton-Raphson algorithm and BFGS (Broyden-Fletcher-Goldfarb-Shanno algorithm) are most often used algorithms [1].

Estimation and its difficulties are described by Terasvirta in [3].

As it is admitted in [1], the numerical optimization is more stable if the transition function is standardized before optimization: in the case of LSTAR1, γ should be divided by the sample standard deviation, but for ESTAR and LSTAR2 – by the sample variance of the variable.

Terasvirta emphasized in [7] that even if convergence is achieved, the model's validity still needs to be analysed. Due to the presence of local minima, especially in relatively short time series, it is important to verify that the obtained estimates are reasonable (e.g., if the threshold values fall within the range of the series and the two thresholds differ). Not always such tests are implemented in software estimation functions. Additionally, it is essential to examine the residuals and their autocorrelation. In general, a more parsimonious model is preferred. If some coefficients (except of γ) of the model appear nonsignificant, it is better to exclude at least part of them from the model.

In popular R software package tsDyn for nonlinear time series models with regime switching, estimation functions only for SETAR and LSTAR1 are offered without a chance to estimate ESTAR and LSTAR2. Since R is an open-code software, the code of function lstar() for LSTAR1 estimation, which already allows for different equation orders, can be used as a base for new functions for other models estimation such as LSTAR2 and ESTAR. It was done by the author of this paper in [8] for LSTAR2.

III. Description of Realization Of the Model Estimation in R

A. *lstar2() function construction*

Function lstar(), available in R package thDyn and used for estimation of LSTAR1 model, was modified into lstar2() function for LSTAR2 model estimation.

The modification was performed using available lstar() function code in R documentation by making the necessary adjustments specifically for the LSTAR2 model. These adjustments were the following: adding two threshold coefficients; modifying the function code by applying quadratic logistic transition function which is using previously mentioned two threshold values *th1* and *th2* as function variables beside smoothing parameter *γ* and threshold variable z_t; also, there were modifications done for internal lstar() function gradEhat() adjusting it for LSTAR2 model by creating derivatives including three variables *γ*, *th1* and *th2* to create Jacobian matrix which is used for coefficients α_i and β_i optimization.

B. *lstar2() function additions*

A few additional improvements were done while creating the lstar2() function. The option for the user to choose from two types of LSTAR2 model forms is implemented, what can be useful for interpretation of results, Namely, the transition function can be used only for 'high' regime regression in form of multiplier G2(z_t, *th1, th2, γ*) or one can use transition function also as multiplier for 'low' regime regression but in form of 1- G2(z_t, *th1, th2, γ*). For more possibilities for user to define starting values for transition function variables, an option to choose starting values for part of the variables was added, allowing the user to define all 3 variables, define *γ*, allowing for the function to choose the threshold coefficients by grid search, or choose starting values for threshold *th1* and *th2*, allowing for function to choose *γ* by grid search.

To analyse estimated model results, summary, predictions and to perform a comparison with other models available in R environment, a few more modifications were done by adjusting tsDyn:::predict.nlar(), tsDyn:::print.lstar() and tsDyn:::print.summary.lstar() for LSTAR2 model estimated by lstar2().

C. *LSTAR2 estimation*

A new function lstar2() is created for efficient LSTAR2 model estimation implementing the mentioned modifications and additions to lstar(). The code is available in [8].

LSTAR2 estimation is performed similar to LSTAR1 model estimation by using concentrated least squares for coefficient estimation. This function definition in R environment for user looks like lstar2(x, m, d, steps, series, mL, mH, mTh, thDelay, thVar, th1, th2, gamma, trace, include, control, starting.control, tp = 1) where the arguments are similar to lstar() function except of possible initial values of two thresholds *th1*, *th2* and the type of model form *tp*.

IV. Data Analysis

As in most other countries, the unemployment rate in Latvia, like many other economic variables, is highly unstable. Moreover, we may expect that it changes differently in different periods being influenced by different economic and political shocks.

A. The data

Our analysis is based on publicly available monthly data of the unemployment rate among 15-75 years old residents of Latvia for the period from January 2002 to June 2024, published by the Central Statistical Bureau of Latvia [9].

Fig. 1. Unemployment rate in Latvia(2002M1-2024M6).

Looking at the shape of our data on Fig.1, we notice that it is strongly nonstationary, with breaks, most likely several. Especially during the Great Financial Crisis of 2008, the unemployment rate rose sharply. Then it was gradually decreasing until the Covid-19 pandemic, when it sharply rose again, although not so strongly as in 2008. After a small period of gradual decrease until spring of 2023, we observe the signs of an increase again.

Such heterogenous dynamics may require a threshold model. It is also confirmed with statistical tests. Dickey-Fuller Augmented test cannot reject the unit root. But Zivot-Andrews test rejects the unit root and accepts the alternative hypothesis of stationary series with a break at an unknown point in either the intercept, the linear trend or in both.

To get rid of a highly expressed trend, we turn to differences (see them at Fig.2) and further analyse the changes of the unemployment rate and construct the model for them.

Now Dickey-Fuller Augmented test rejects the unit root with the p-value of 0.03624. Although we also do not see a strong evidence of nonstationarity in the correlogram (autocorrelation and partial autocorrelation functions) (Fig.3), at least in first 18 lags, Tsay's test rejects the null hypothesis that the time series follows some autoregressive process and accepts the alternative hypothesis about nonlinearity with p-value close to zero (1.413222e-07).

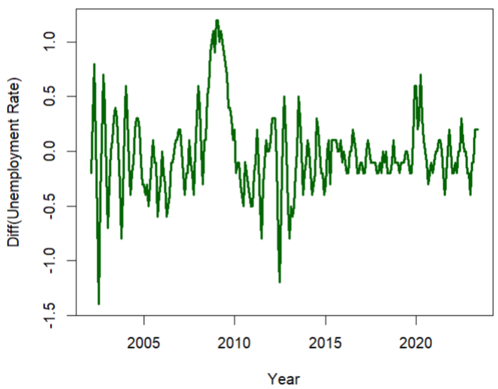

Fig. 2. Differences of Unemployment rate in Latvia (2002M2-2024M6).

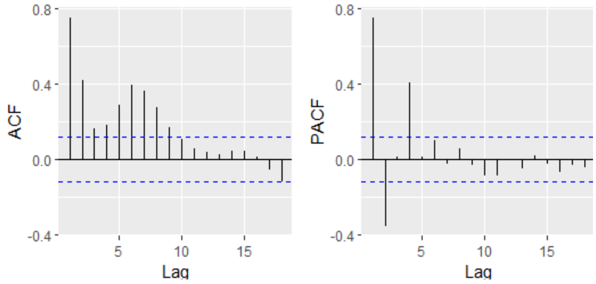

Fig. 3. Autocorrelation (ACF) and Partial autokorrelation function (PACF) of the differences of Unemployment rate

Zivot-Andrews test rejects the unit root also in the series of differences and accept alternative hypothesis of stationary series with a break at an unknown point. Then, Threshold nonlinearity test rejects H_0 of being an autoregressive process and accepts H_1 of being threshold autoregressive model for all lags at least till 12 with p-values close to zero. Also, the special Terasvirta (1994) nonlinearity test is applied to test nonlinearity and to choose among the smooth transition models. Testing regression orders until 12, linearity is rejected for them all, and first order LSTAR model is chosen in all cases.

As follows from the tests, a threshold or smooth transition autoregressive model can describe our data better than linear autoregressive models.

B. Model choise and comparison

Henceforth, the period up to June 2023 is used as a training period estimating the models and the last 12 months are used as a testing period.

Despite the majority of the performed tests point to LSTAR as the model to most likely have the best fit, ARIMA, SARIMA, SETAR, LSTAR1, LSTAR2 of the best possible order and parameters with good properties (full, stationary, parsimonious) are estimated for the comparison. All types models of order less than 4 are not full. Therefore, the following models are chosen:

AR (4)

$$X_t = 1.0383X_{t-1} - 0.2164X_{t-2} - 0.4507X_{t-3} + 0.4425\,X_{t-4} + w_t$$

SARIMA(4,0,0)(1,0,0)[12]

$$(1 - 0.0505B^{12})(1 - 1.043\text{B} + 0.2211B^2 + 0.4507B^3 - 0.4391B^4)X_t = w_t$$

SETAR

$$X_t = \begin{cases} -0.0592 + 0.8059X_{t-1} - 0.1566X_{t-2} - 0.5272X_{t-3} + 0.3105X_{t-4} + w_t^{(1)}, & if \quad X_{t-1} \le 0.2 \\ -0.0471 + 1.1059X_{t-1} - 0.4237X_{t-2} + 0.1746X_{t-3} + 0.1424X_{t-4} + w_t^{(2)}, & if \quad X_{t-1} > 0.2 \end{cases}$$

LSTAR1

$$X_t = -0.034 + 0.8987X_{t-1} - 0.2087X_{t-2} - 0.4427X_{t-3} + 0.3195X_{t-4} +$$

$$+ \frac{1}{1+e^{-11.9921(X_{t-1}-0.7322)}}[0.506 - 0.4994X_{t-1} + 0.4883X_{t-2}] + w_t,$$

Despite LSTAR was recommended by Terasvirta test, LSTAR2 also was estimated for comparison.

LSTAR2

$$X_t = -0.0314 + 0.9803X_{t-1} - 0.2218X_{t-2} - 0.4349X_{t-3} + 0.3047X_{t-4}$$

$$+ \left(\frac{1}{1+e^{-9.8589(y_{t-1}+1.1573)(y_{t-1}-0.7911)}}\right)[0.4263 - 0.3467X_{t-1} + 0.34874X_{t-2}] + w_t,$$

Comparing the fit of the chosen SARIMA, AR(4), LSTAR1, SETAR, and models using the root mean squared error loss (RMSE), LSTAR2 displays the best goodness of fit (Table I).

TABLE I. COMPARISON OF MODELS WITH RMSE

Model	RMSE	Goodness of fit
AR(4)	0.1982587	
SARIMA(4,0,0)(1,0,0)[12]	0.1980783	
SETAR	0.1846241	
LSTAR1	0.1889631	
LSTAR2	**0.1830187**	**Fits the best**

C. *Analysis of the best model*

The majority of the coefficients of the estimated LSTAR2 model are statistically significant (Table II).

Non-linearity test rejects linear autoregressive model and accepts logistic smooth transition model. The thresholds estimated as -1.1573 and 0.7911. So -0.1831 is the level mostly corresponding to the second regime. But the mentioned thresholds are the levels when transition starts and ends. The smoothing coefficient is 9.8589 which is moderate, nor large nor small, what means that the transition is not too fast, but also it is not slow.

The chosen LSTAR2 model is asymmetric, i.e. the equations of the regimes have different orders, what lead to a more parsimonious model. Some coefficients remained unsignificant because their exclusion from the model lead to nonstationarity or the estimation procedure did not converge.

TABLE II. ESTIMATES OF THE COEFFICIENTS OF THE CHOSEN LSTAR2 MODEL

Coefficient	Estimate	Std. Error	t value	Pr(>|z|)
const.L	-0.031432	0.012447	-2.5253	0.011561
phiL.1	0.980343	0.059585	16.4528	< 2.2e-16
phiL.2	-0.221793	0.074143	-2.9914	0.002777
phiL.3	-0.434928	0.073643	-5.9059	3.507e-09
phiL.4	0.304699	0.054349	5.6064	2.066e-08
const.H	0.426263	0.144167	2.9567	0.003109
phiH.1	-0.346730	0.251428	-1.3790	0.167882
phiH.2	0.348695	0.360073	0.9684	0.332843
gamma	9.858945	19.145107	0.5150	0.606582
th1	-1.157334	0.034221	-33.8197	< 2.2e-16
th2	0.791095	0.032642	24.2352	< 2.2e-16
Non-linearity test of full-order LSTAR model against full-order AR model F = 9.8649 ; p-value = 2.0536e-07				

According to Diebold-Mariano test for predictive accuracy of the constructed LSTAR1 and LSTAR2 for in-sample one-step forecasts, in the sample period until June 2023 one can reject the null hypothesis that the two models have the same forecast accuracy and accept the alternative that LSTAR2 forecast is more accurate with p-value=0.05405.

From Fig.4 one can see that LSTAR1 un LSTAR2 forecasts for 12 steps, i.e. for the period July 2023 – June 2024 based on the previous data, are very close. The equations of these models are of the same order but their smoothing functions have different structure. ARIMA, SARIMA forecasts are too high, SETAR forecast is too low.

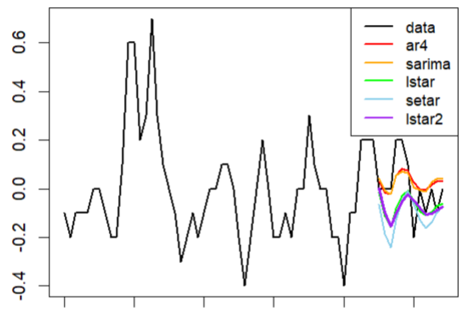

Fig. 4. The forecast of the difference of unemployment by different models.

Predicting the difference of the rate of unemployment for the next 12 month based on the all available data up to June 2024, we see from Fig.5 that LSTAR2 and LSTAR1 forecasts are close too. The difference is predicted negative what means the decrease of the unemployment rate.

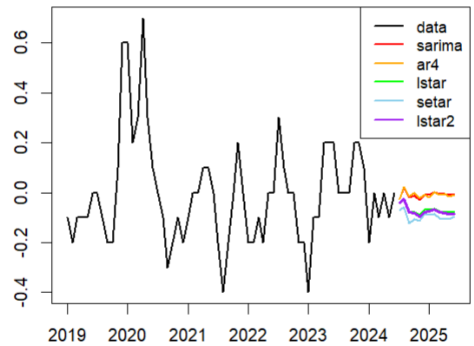

Fig. 5. The forecast of the future values of the difference of unemployment by different models.

V. CONCLUSION

The differences of the unemployment rate in Latvia were explored with various tests whether they have structural brakes, and follow threshold or a smooth transition model.

The second order logistic smooth transition autoregressive model is found to have a better fit and forecast for these series than the linear autoregression and multiplicative seasonal autoregressive model, better than self-exciting threshold autoregression and smooth transition autoregressive model with transition function of the first order.

The estimation function for LSTAR2 model developed in [8] is useful for the modelling and forecasting of economic and financial data with changes in their structure. The possibility to use different orders for the regimes' equations allows to use more parsimonious models.

VI. FUTURE RESEARCH DIRECTIONS

The accuracy of the estimation procedure should be more tested and improved if necessary. Also, it would be useful to develop an estimation function in R for exponential smooth transition autoregressive model that may be used for changing economic data. And finally, the new estimation functions for different smooth transition models should be combined into the new R package and made available for public.

REFERENCES

[1] Kavkler A., Mikek P., Böhm B., Boršič D. Nonlinear econometric models: the smooth transition regression approach [online]. 2008 [viewed 5 May 2024]. Available from: https://www.researchgate.net/publication/228556576

[2] Aydin D, Mermi S. Some regime-switching models for economic time series: a comparative study. Eskisehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering. 2022, 23(1), pp.48-69, DOI: 10.18038/estubtda.881251

[3] Hansen, G. Forecasting prices of dairy commodities – a comparison of linear and nonlinear models. Irish Journal of Agricultural and Food Research. 59(1) 2020 DOI: 10.15212

[4] Pavļenko, O., Matvejevs, A. Comparison of linear and nonlinear models for forecasting of food commodity prices in Latvia. In: 22nd International Scientific Conference "Engineering for Rural Development" : Proceedings, Vol.22, Latvia, Jelgava, 24-26 May, 2023. Jelgava: Latvia University of Life Sciences and Technologies, 2023, pp.735-744. ISSN 1691-5976. Available from: doi:10.22616/ERDev.2023.22.TF145

[5] Enders, W. Applied Econometric Time Series. Fourth edition, Publisher: Wiley; 4 edition (October 22, 2014) ISBN-13: 978-1118808566 ISBN-10: 1118808568

[6] Robert H. Shumway, David S. Stoffer Time Series Analysis and Its Applications: With R Examples. Third or Fourth edition Springer Science + Business Media, LLC 2011, 2017.

[7] Terasvirta T., Specification, Estimation, and Evaluation of Smooth Transition Autoregresive Models. Journal of the Am.St.As.,1994

[8] Zeltina L. Loģistiskās gludās pārejas autoregresīvā modeļa lstar2 novērtēšanas implementācija rstudio un pielietojumi finansēs. Maģistra darbs. Rīga: RTU, 2024.

[9] The data of unemployed and unemployment rate aged 15-74 by sex 2002M01-2024M06 [Online]. Available: https://data.stat.gov.lv/pxweb/lv/OSP_PUB/START__EMP__NBBA__NBBB/NBB150m/