Forecasting Unemployment Rate in Lubbock County

Eco 4306 Economic and Business Forecasting

Spring 2019

Introduction

Goal

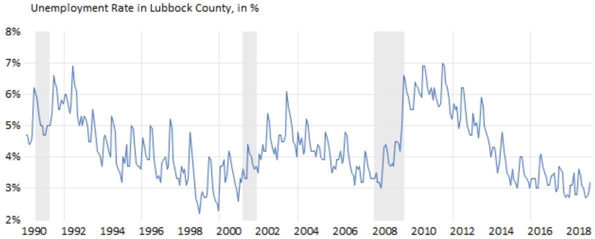
- present a seasonal ARMA model suitable to forecast monthly unemployment rate in Lubbock County, Texas

Outline

- Data
- Seasonal ARMA model
  - Estimation
  - Forecast
  - Forecast Evaluation and Comparison with Naive Forecasting Method
- Conclusion

Data

- monthly data for the Unemployment Rate in Lubbock County, TX
- obtained from FRED database, see code TXLUBB3URN
- sample: January 1990 to January 2019



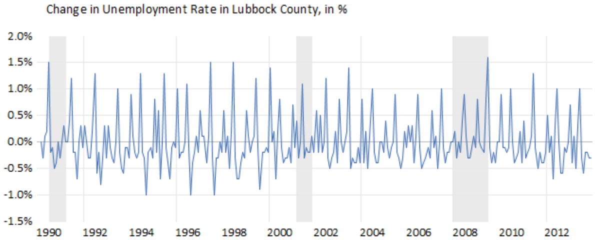
- estimation sample: January 1990 to December 2013
- prediction sample: January 2014 to January 2019

Data

- first difference applied to obtain the change in the unemployment rate

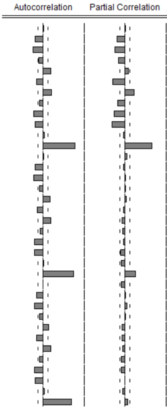
$$y_t = \Delta UR_t = UR_t - UR_{t-1}$$

- time series y_t exhibits seasonal variation



Data

- correlogram for change in unemployment rate $y_t = \Delta UR_t$ confirms the presence of a seasonal pattern
- large spike in PAC at lags 12 and 24, and large spikes at multiples of 12 in AC



	AC	PAC	Q-Stat	Prob
1	0.011	0.011	0.0320	0.858
2	-0.200	-0.200	11.643	0.003
3	-0.246	-0.252	29.378	0.000
4	-0.104	-0.167	32.542	0.000
5	0.189	0.086	43.103	0.000
6	-0.167	-0.309	51.339	0.000
7	0.212	0.233	64.702	0.000
8	-0.088	-0.186	66.986	0.000
9	-0.231	-0.263	82.869	0.000
10	-0.204	-0.328	95.356	0.000
11	0.031	-0.046	95.637	0.000
12	0.816	0.693	296.39	0.000
13	0.021	0.061	296.52	0.000
14	-0.193	-0.042	307.87	0.000
15	-0.240	0.010	325.40	0.000
16	-0.098	0.009	328.32	0.000
17	0.175	0.039	337.73	0.000
18	-0.153	-0.029	344.99	0.000
19	0.197	-0.069	356.96	0.000
20	-0.077	-0.050	358.81	0.000
21	-0.220	-0.034	373.89	0.000
22	-0.218	-0.101	388.71	0.000
23	0.022	-0.092	388.86	0.000
24	0.777	0.274	579.28	0.000
25	-0.002	-0.085	579.28	0.000
26	-0.171	0.014	588.63	0.000
27	-0.221	0.059	604.18	0.000
28	-0.088	0.018	606.65	0.000
29	0.142	-0.069	613.14	0.000
30	-0.171	-0.074	622.54	0.000
31	0.194	-0.064	634.73	0.000
32	-0.088	-0.098	637.22	0.000
33	-0.216	-0.083	652.48	0.000
34	-0.201	-0.004	665.67	0.000
35	0.026	-0.061	665.89	0.000
36	0.725	0.067	839.75	0.000

Estimated AR(1)-SARMA(1,1) Model

- multiplicative seasonal AR(1)-SARMA(1,1) model

$$(1 - \phi_1 L)(1 - \phi_{12} L^{12}) y_t = \phi_0 + (1 + \theta_{12} L^{12}) \varepsilon_t$$

- estimation results

Dependent Variable: D(TXLUBB3URN)
Method: ARMA Maximum Likelihood (BFGS)
Date: 03/28/19 Time: 09:55
Sample: 1990M02 2013M12
Included observations: 287
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004496	0.172417	-0.026075	0.9792
AR(1)	-0.098539	0.056171	-1.754261	0.0805
SAR(12)	0.994938	0.002843	350.0173	0.0000
MA(12)	-0.788612	0.042420	-18.59037	0.0000
SIGMASQ	0.046547	0.003136	14.84395	0.0000

R-squared	0.807475	Mean dependent var	-0.002439
Adjusted R-squared	0.804744	S.D. dependent var	0.492561
S.E. of regression	0.217652	Akaike info criterion	-0.090124
Sum squared resid	13.35900	Schwarz criterion	-0.026370
Log likelihood	17.93274	Hannan-Quinn criter.	-0.064572
F-statistic	295.6857	Durbin-Watson stat	1.987430
Prob(F-statistic)	0.000000		

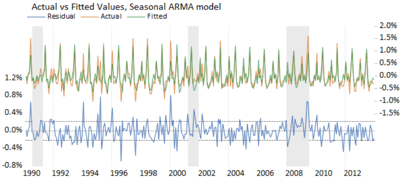
- estimated model thus takes the form

$$(1 + 0.098L)(1 - 0.995L^{12}) y_t = -0.004 + (1 - 0.789L^{12}) \varepsilon_t$$

In-Sample Evaluation (Checking Model for Adequacy)

residuals of the estimated AR(1)-SARMA(1,1) model appear to be white noise

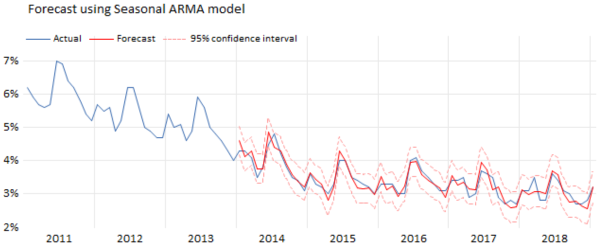
- time series plot does not show any recognizable pattern, or any changes in volatility over the estimation sample
- no remaining significant time dependence in correlogram



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.005	0.005	0.0071	
		2	0.024	0.024	0.1816	
		3	-0.024	-0.025	0.3555	
		4	-0.043	-0.043	0.8941	0.344
		5	-0.014	-0.013	0.9544	0.621
		6	0.024	0.025	1.1182	0.773
		7	0.107	0.106	4.5139	0.341
		8	0.057	0.054	5.4963	0.358
		9	0.022	0.016	5.6352	0.465
		10	0.015	0.019	5.7016	0.575
		11	0.016	0.027	5.7755	0.672
		12	0.064	0.073	7.0297	0.634
		13	0.035	0.035	7.4016	0.687
		14	0.023	0.009	7.5590	0.752
		15	-0.018	-0.028	7.6613	0.811
		16	-0.062	-0.065	8.8513	0.784
		17	-0.015	-0.016	8.9175	0.836
		18	0.093	0.090	11.596	0.709
		19	-0.025	-0.047	11.794	0.758
		20	0.096	0.069	14.630	0.622
		21	-0.029	-0.039	14.898	0.669
		22	-0.095	-0.097	17.735	0.540
		23	-0.028	-0.011	17.976	0.589
		24	0.050	0.064	18.776	0.600
		25	-0.097	-0.123	21.747	0.475
		26	0.045	0.029	22.395	0.497
		27	0.047	0.040	23.105	0.514
		28	0.003	0.008	23.108	0.571
		29	-0.135	-0.125	28.935	0.314
		30	-0.033	-0.027	29.283	0.347
		31	-0.081	-0.084	31.403	0.299
		32	-0.020	-0.019	31.527	0.341
		33	-0.037	-0.040	31.969	0.369
		34	-0.037	-0.036	32.415	0.397
		35	0.010	0.008	32.447	0.445
		36	-0.002	0.022	32.448	0.494

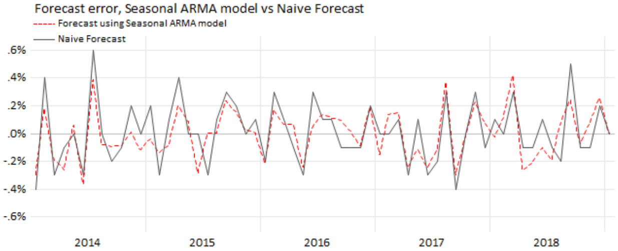
Forecast based on AR(1)-SARMA(1,1) model

- sequence of one step ahead forecasts
- prediction sample January 2014 to January 2019
- forecast tracks actual data quite well



Forecast Evaluation - Root Mean Square Error (RMSE)

- simple naive forecast for the change in unemployment rate $f_{t,1}^{naive} = y_{t+1-12}$
- implied naive forecast for unemployment rate $\widehat{UR}_{t,1}^{naive} = UR_t + f_{t,1}^{naive}$
- forecast errors $e_{t+1} = y_{t+1} - f_{t,1}$



- root mean squared error (RMSE)
  0.180 for forecast based on the AR(1)-SARMA(1,1) model
  0.221 for naive forecast

Forecast Evaluation - Equal Predictive Ability Test

- test whether the difference in the precision of the two forecasts is statistically significant
- hypothesis $H_0 : \beta_0 = 0$ for the regression

$$\Delta L_{t,1} = \beta_0 + u_t$$

where

$$L_{t,1} = L(e_{t,1}^{SARMA}) - L(e_{t,1}^{naive})$$

is the difference between the losses associated with the two alternative forecasts

Dependent Variable: DL_NAIVE
Method: Least Squares
Date: 03/28/19 Time: 09:55
Sample: 2014M01 2019M01
Included observations: 61

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.016630	0.006869	-2.421098	0.0185

R-squared	0.000000	Mean dependent var	-0.016630
Adjusted R-squared	0.000000	S.D. dependent var	0.053645
S.E. of regression	0.053645	Akaike info criterion	-2.996584
Sum squared resid	0.172670	Schwarz criterion	-2.961979
Log likelihood	92.39581	Hannan-Quinn criter.	-2.983022
Durbin-Watson stat	2.019540		

- since p-value for $\hat{\beta}_0$ is 0.0185 difference is indeed statistically significant at 5% level
- AR(1)-SARMA(1,1) thus produced a more precise forecast than the naive method

Conclusion

- data for unemployment rate in Lubbock County is only available since 1990
- sample is thus relatively short
- seasonal ARMA model however performs quite well when applied to create the step ahead forecast for the unemployment rate
- estimated model outperforms the naive forecasting method, producing significantly more precise forecast