Forecasting Unemployment Rate in Lubbock County

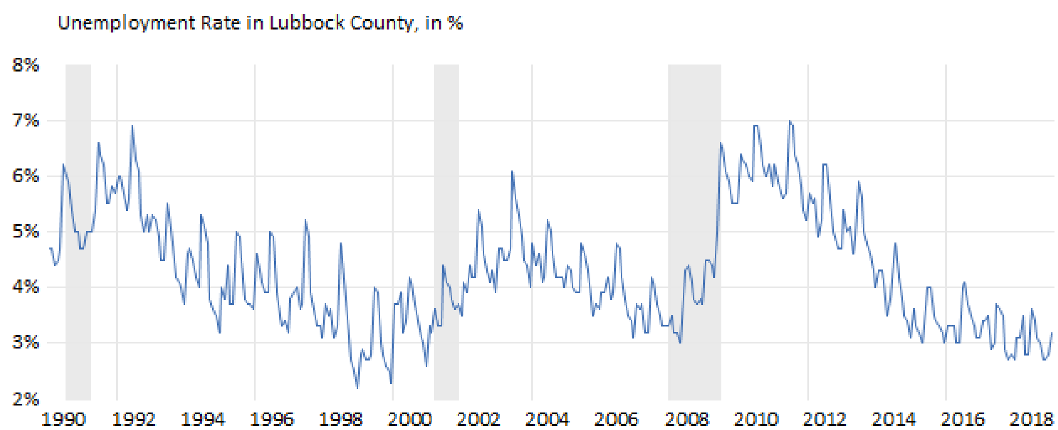
Eco 4306 Economic and Business Forecasting
Spring 2019

Introduction

The goal of this short report is to present the analysis undertaken to estimate a seasonal ARMA model in order to be able to forecast monthly unemployment rate in Lubbock County, Texas. First, the data is briefly introduced. Next, the estimated model is presented and discussed. Finally, precision of the forecast using the seasonal ARMA model is compared to the precision of the naive forecasting method.

Data

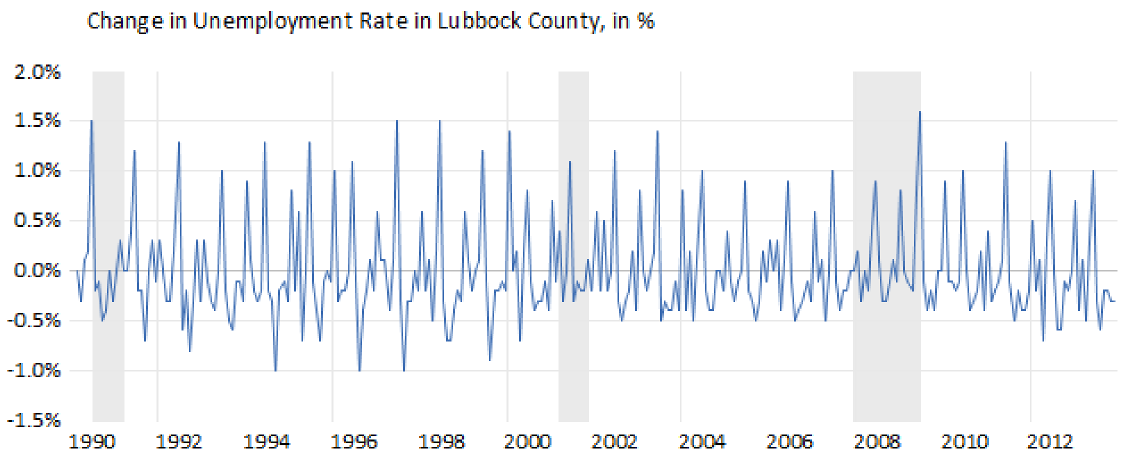
Monthly data for the Unemployment Rate in Lubbock County, TX for the period from January 1990 to January 2019 was obtained from FRED, where it is available under the code TXLUBB3URN. Figure below shows the time series plot for the whole sample.



The estimation sample used to identify and estimate suitable model is January 1990 to December 2013. The remaining part, from January 2014 to January 2019 will be used to evaluate the precision of the forecast. To address the potential problem with non-stationarity, first difference was applied to obtain the change in the unemployment rate

$$y_t = \Delta TXLUBB3URN_t = TXLUBB3URN_t - TXLUBB3URN_{t-1}$$

Figure below shows this transformed time series y_t, which exhibits seasonal variation with significant spikes in January each year when less work is available in agriculture and construction, and also in June of each year when new graduate enter the labor market.



The correlogram of the change in the unemployment rate y_t shown in Appendix A confirms the presence of a seasonal pattern, reflected by a large spike in PAC at lags 12 and 24, and large spikes at multiples of 12 in AC.

Model Estimation

To account for the seasonal pattern (large spike in PAC at lags 12 and 24, and large spikes at multiples of 12 in AC) and also the significant non-seasonal time dependence (significant lags 2 to 4 and 6 to 10 in PAC), a multiplicative seasonal AR(1)-SARMA(1,1) model was estimated

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

Figure below shows the results of the estimation

Dependent Variable: D(TXLUBB3URN)
Method: ARMA Maximum Likelihood (BFGS)
Date: 03/28/19 Time: 09:24
Sample: 1990M02 2013M12
Included observations: 287
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients

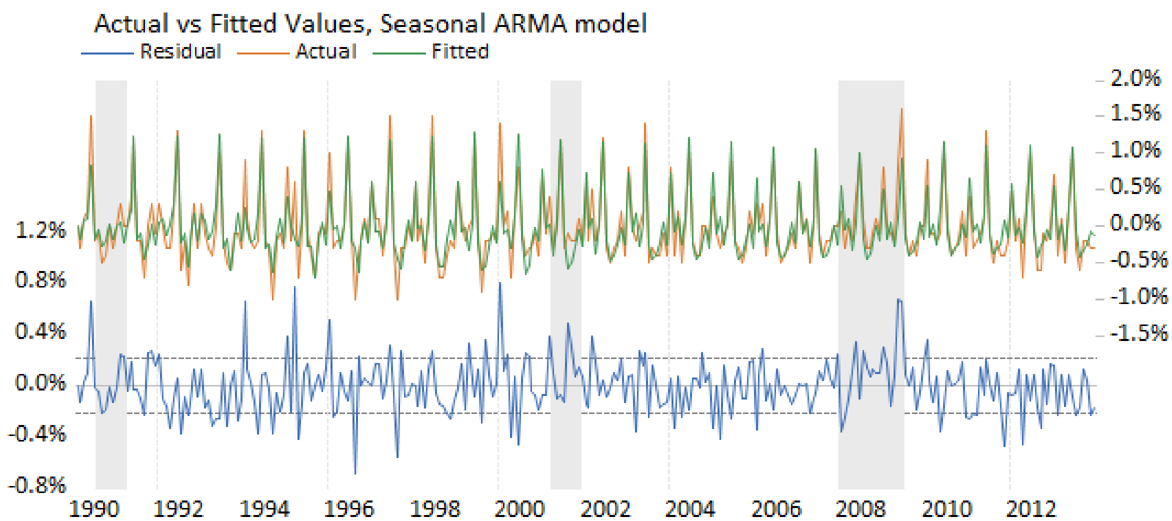
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004496	0.172417	-0.026075	0.9792
AR(1)	-0.098539	0.056171	-1.754261	0.0805
SAR(12)	0.994938	0.002843	350.0173	0.0000
MA(12)	-0.788612	0.042420	-18.59037	0.0000
SIGMASQ	0.046547	0.003136	14.84395	0.0000

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		

The estimated model thus takes the form

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

Residuals are plotted below, they do not appear to show any recognizable pattern, or any changes in volatility over the estimation sample.

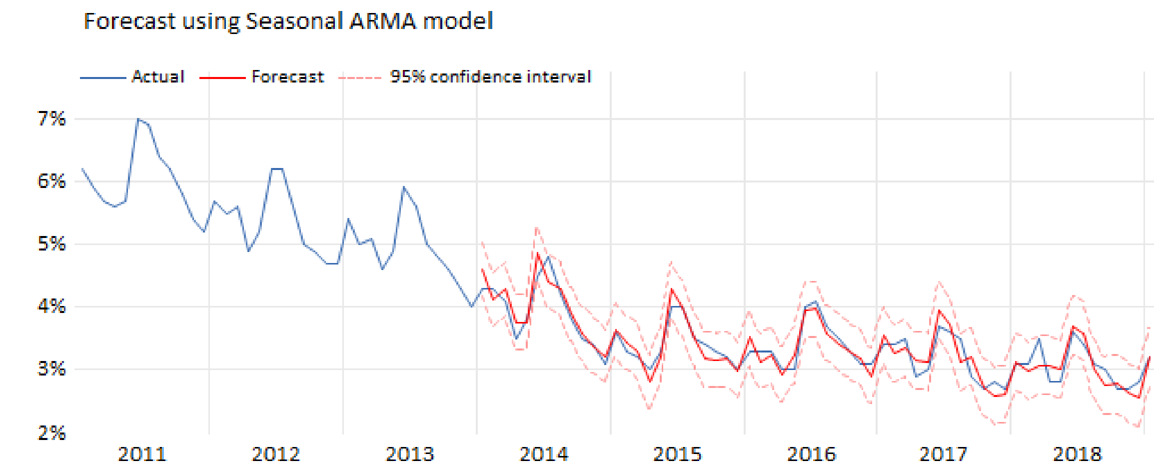


Residuals also do not appear to show any remaining significant time dependence, as shown in their correlogram in Appendix B. All components of the AC and PAC functions are within the 95% confidence interval around 0 and the p-values of Ljung-Box statistic are in general large, and are all above 0.3. We can thus conclude that the residuals of the estimated AR(1)-SARMA(1,1) model appear to be white noise.

Forecast

As mentioned above, period from January 2014 to January 2019 was used as prediction sample to evaluate the precision of the forecast obtained using the estimated AR(1)-SARMA(1,1) model.

First, a sequence of one step ahead forecasts was created, together with their 95% confidence interval. They are plotted in the figure below. In general the model forecast is reasonably precise, and can fit the actual path of unemployment rate in Lubbock county well.



Next, as a benchmark, a simple naive forecast was created for the change in unemployment rate in the same prediction sample January 2014 to January 2019 using

$$f_{t,1}^{naive} = y_{t+1-12}$$

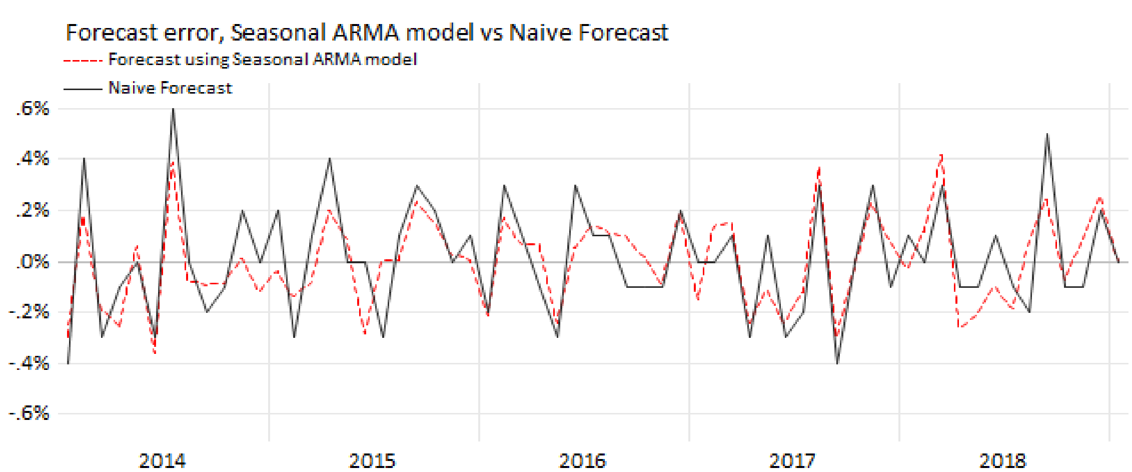
This essentially means that the change in the unemployment rate is predicted to be the same as it was in the same month one year ago. The implied forecast for unemployment rate $TXLUBB3URN_{t,1}^{naive}$ was afterwards calculated by adding the forecasted change in unemployment rate in the next month $f_{t,1}^{naive}$ to the actual unemployment rate in the current month $TXLUBB3URN_t$, that is, using

$$\widehat{TXLUBB3URN}_{t,1}^{naive} = TXLUBB3URN_t + f_{t,1}^{naive}$$

For both forecasts the forecast errors are calculated as

$$e_{t+1} = y_{t+1} - f_{t,1}$$

and are plotted below



The forecast errors tend to be somewhat smaller in the magnitude. The root mean squared error (RMSE) for forecast using the AR(1)-SARMA(1,1) model is 0.180, while the root mean squared error for forecast using naive forecasting method is 0.221. Thus for both methods it is roughly 0.2 percentage points in the sample where the unemployment rate has been fluctuating in the range from 2.8 to 4.8 percentage points.

To determine whether the difference in the precision between the two forecasts is statistically significant or not, the test for the equal predictive ability was performed. This was done by estimating a simple regression model

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

where $L_{t,1}$ is the difference between the losses associated with the two alternative forecasts

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

and testing the hypothesis $H_0 : \beta_0 = 0$. Rejecting this hypothesis means that the two model do not have equal predictive power. The results for the estimated regression show that the difference is indeed statistically significant at 5% level, since p-value for $\hat{\beta}_0$ is 0.0185. The AR(1)-SARMA(1,1) model thus produced a more precise forecast than the naive forecast.

Dependent Variable: DL_NAIVE
Method: Least Squares
Date: 03/28/19 Time: 09:24
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		

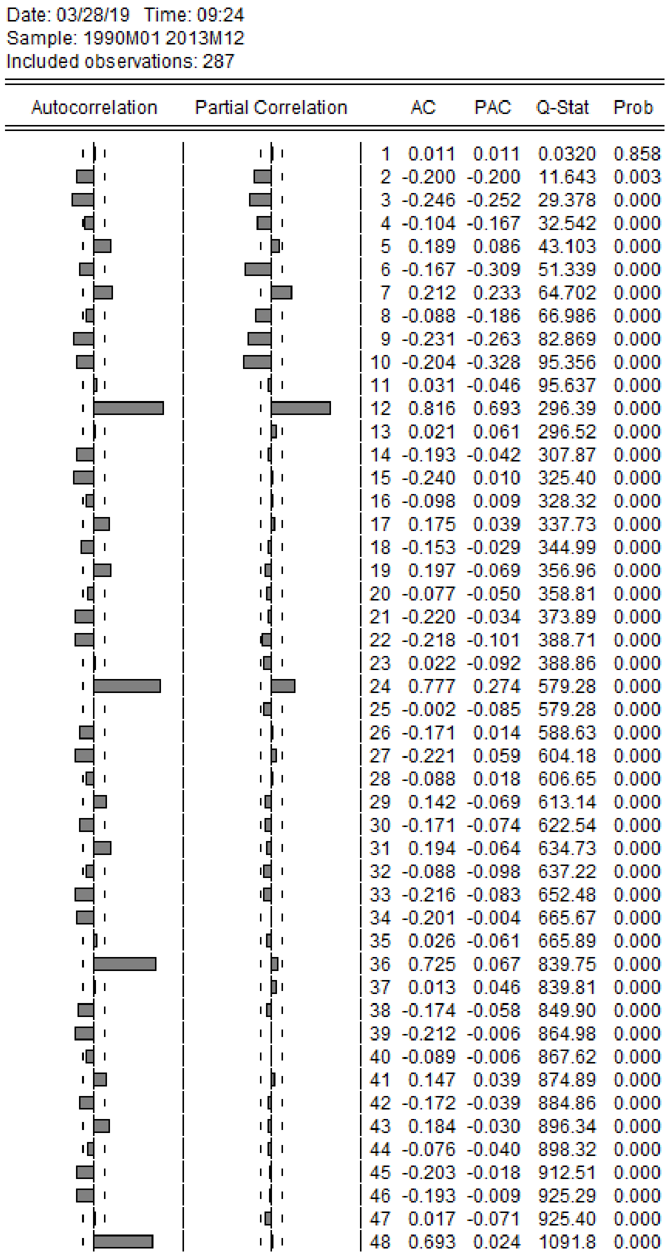
Conclusion

As shown in this short note, even thought the data for unemployment rate in Lubbock County is only available since 1990, and thus the sample is relatively short, seasonal ARMA model performs quite well when applied to create its one step ahead forecast. The estimated model outperforms the naive forecasting method, producing significantly more precise forecast.

Appendix A

Figure below shows the correlogram for the first difference in the unemployment rate

$$y_t = \Delta TXLUBB3URN_t = TXLUBB3URN_t - TXLUBB3URN_{t-1}$$



Date: 03/28/19 Time: 09:24
Sample: 1990M01 2013M12
Included observations: 287

	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
37	0.013	0.046	839.81	0.000
38	-0.174	-0.058	849.90	0.000
39	-0.212	-0.006	864.98	0.000
40	-0.089	-0.006	867.62	0.000
41	0.147	0.039	874.89	0.000
42	-0.172	-0.039	884.86	0.000
43	0.184	-0.030	896.34	0.000
44	-0.076	-0.040	898.32	0.000
45	-0.203	-0.018	912.51	0.000
46	-0.193	-0.009	925.29	0.000
47	0.017	-0.071	925.40	0.000
48	0.693	0.024	1091.8	0.000

Appendix B

Figure below shows the correlogram for the residuals from the estimated AR(1)-SARMA(1,1) model

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

Date: 03/28/19 Time: 09:24
Sample: 1990M01 2013M12
Included observations: 287
Q-statistic probabilities adjusted for 3 ARMA terms

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
		37	-0.032	-0.009	32.786	0.527
		38	0.016	0.009	32.873	0.571
		39	-0.019	0.008	32.999	0.612
		40	-0.043	-0.019	33.620	0.628
		41	-0.044	-0.026	34.266	0.643
		42	-0.062	-0.038	35.562	0.627
		43	-0.083	-0.043	37.916	0.564
		44	0.043	0.022	38.554	0.580
		45	-0.007	-0.003	38.572	0.622
		46	-0.046	-0.064	39.299	0.633
		47	-0.032	-0.042	39.658	0.658
		48	0.014	0.025	39.731	0.694

Appendix C

Below is the EViews code used.

```
series dtxlubb3urn = txlubb3urn - txlubb3urn(-1)

' time series plot - txlubb3urn, whole sample
freeze(g_txlubb3urn_ts_whole) txlubb3urn.line
g_txlubb3urn_ts_whole.recshade
g_txlubb3urn_ts_whole.addtext(ac) ""
g_txlubb3urn_ts_whole.addtext(al)  "Unemployment Rate in Lubbock County, in %"

' estimation sample
smpl 1990m1 2013m12

' time series plot - dtxlubb3urn
freeze(g_dtxlubb3urn_ts) dtxlubb3urn.line
g_dtxlubb3urn_ts.recshade
g_dtxlubb3urn_ts.addtext(ac) ""
g_dtxlubb3urn_ts.addtext(al) "Change in Unemployment Rate in Lubbock County, in %"

' correlogram - dtxlubb3urn
freeze(g_dtxlubb3urn_corr) dtxlubb3urn.correl(48)

' estimate seasonal AR-SARMA model
equation eq1.ls d(txlubb3urn) c ar(1) sar(12) sma(12)
freeze(tbl_eq1) eq1

' time series plot - residuals
freeze(g_eq1_resid_ts) eq1.resid
g_eq1_resid_ts.recshade
g_eq1_resid_ts.legend position(0,-0.35)
g_eq1_resid_ts.legend -inbox
g_eq1_resid_ts.addtext(al) "Actual vs Fitted Values, Seasonal ARMA model"

' correlogram - residuals
freeze(g_eq1_resid_corr) eq1.correl(48)

' fixed scheme forecast and its 95% confidence interval
smpl 2014m1 2019m2
freeze(tbl_eq1_f_fixed) eq1.fit(f=na, e, g) txlubb3urn_f @se txlubb3urn_f_se

series txlubb3urn_f_lb = txlubb3urn_f - 1.96* txlubb3urn_f_se
series txlubb3urn_f_ub = txlubb3urn_f + 1.96* txlubb3urn_f_se
```

```
' naive forecast
series dtxlubb3urn_f_naive = dtxlubb3urn(-12)
series txlubb3urn_f_naive = txlubb3urn(-1) + dtxlubb3urn_f_naive

' plot forecast
smpl 2014m1 2019m2
smpl 2011m1 2019m2

graph g_eq1_f.line txlubb3urn txlubb3urn_f txlubb3urn_f_lb txlubb3urn_f_ub

g_eq1_f.setelem(1) linecolor(@rgb(0,0,0)) legend("Actual Data")
g_eq1_f.setelem(2) linecolor(@rgb(0,0,255)) legend("Fixed Scheme Forecast")
g_eq1_f.setelem(3) linecolor(@rgb(150,150,255)) legend("95% confidence interval")
g_eq1_f.setelem(4) linecolor(@rgb(150,150,255)) legend("")
g_eq1_f.options linepat
g_eq1_f.legend columns(3)
g_eq1_f.legend position(0,-0.1)
g_eq1_f.legend -inbox
g_eq1_f.addtext(al) "Forecast using Seasonal ARMA model"

' calculate and plot forecast errors
series txlubb3urn_e = txlubb3urn - txlubb3urn_f
series txlubb3urn_e_naive = txlubb3urn - txlubb3urn_f_naive

graph g_eq1_e.line txlubb3urn_e txlubb3urn_e_naive

g_eq1_e.setelem(1) linecolor(@rgb(200,0,0)) legend("Forecast using Seasonal ARMA model")
g_eq1_e.setelem(2) linecolor(@rgb(0,0,200)) legend("Naive Forecast")
g_eq1_e.legend position(0,-0.35)
g_eq1_e.legend -inbox
g_eq1_e.addtext(al)  "Forecast error, Seasonal ARMA model vs Naive Forecast"

' equal predictive ability test
series L_ar = txlubb3urn_e^2
series L_naive = txlubb3urn_e_naive^2
series dL_naive = L_ar - L_naive

equation eq1_dL.ls dL_naive c
freeze(tbl_eq1_dL) eq1_dL
```