

# 여성 경제활동인구에 관한 시계열 분석

2019년 12월

성신여자대학교 통계학과 20171644 정세영

## 1. 서론

### 1.1 연구주제

우리나라는 가부장적인 문화로 인하여 예로부터 여성의 경제 활동을 금해왔다. 그러나 시대가 변화하면서 점차 경제 활동에 참여하는 여성의 수가 증가하였고, 최근에는 여성 인권 신장을 통한 양성 평등을 지향하는 페미니즘이 수면 위로 떠올랐다. 이 때문에 여성 경제활동인구는 여성 인권 신장을 확인하는 주요 지표 중 하나이다. 본 프로젝트에서는 시간이 지남에 따라 증가하는 여성 경제활동인구를 시계열 분석하여 예측해보고자 한다.

### 1.2 자료 설명

분석에 사용한 자료는 1999년 6월부터 2019년 10월까지 '성별 경제활동인구 총괄' 분기별 자료로, 총 81개의 시점으로 이루어져있고 전체 항목에서 여성 경제활동인구만을 추출하여 사용하였다. 여기서 경제활동인구는 만 15세 이상 인구 중 취업자를 말한다. 자료는 국가통계포털로부터 제공받았다.

## 2. 분석 절차

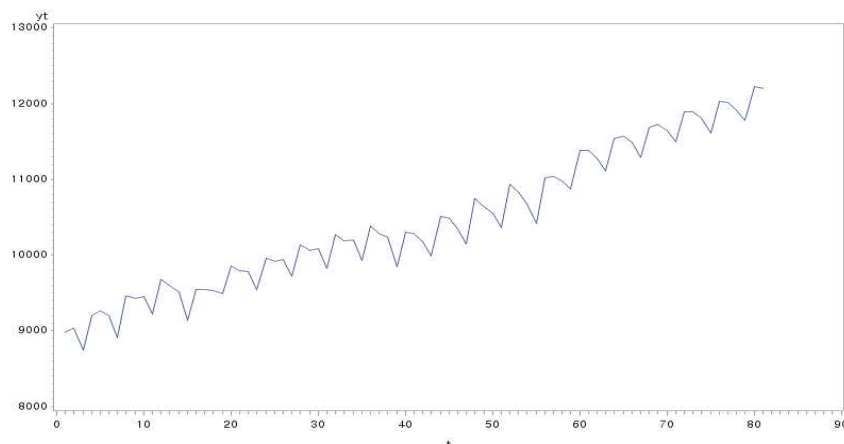
분석은 ARIMA 모형과 분해법 두 가지로 실행한 후 비교하는 절차로 진행하였다.

### 2.1 ARIMA 모형

#### 2.1.1 시도표

가장 먼저 자료의 생김새를 파악하기 위해 시도표를 그려보면 그림 2.1과 같다.

<그림 2.1> 분기별 여성 경제활동인구( $Y_t$ )에 대한 시도표



### 2.1.2 모형 식별

시도표<그림 2.1>을 살펴보면 시간이 경과함에 따라 인구수가 증가하고 있으므로 비정상성을 띠고 있으며, 이런 현상은 <표 2.1>에 있는 SACF가 느리게 감소하는 패턴에서도 볼 수 있다. 따라서 주어진 시계열을 1차 차분하는 것이 바람직하다.

<표 2.1>  $Y_t$ 의 SACF와 SPACF

Name of Variable = yt		
Mean of Working Series		10434.6
Standard Deviation		916.5537
Number of Observations		81

Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	1.00000	*****
1	0.91958	*****
2	0.87058	*****
3	0.83868	*****
4	0.84150	*****
5	0.76631	*****
6	0.72100	*****
7	0.69055	*****
8	0.69339	*****
9	0.62064	*****
10	0.57726	*****
11	0.54952	*****
12	0.55259	*****
13	0.48380	*****
14	0.44188	*****
15	0.41079	*****
16	0.40626	*****
17	0.33532	*****
18	0.29124	*****
19	0.26541	*****
20	0.26663	*****
21	0.20278	*****
22	0.16385	*****
23	0.14007	*****
24	0.14423	*****

Partial Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	0.91958	*****
2	0.16164	*****
3	0.12552	*****
4	0.27065	*****
5	-0.42112	*****
6	0.08377	*****
7	0.06561	*****
8	0.10373	*****
9	-0.26242	*****
10	0.05488	*****
11	0.05473	*****
12	0.03846	*****
13	-0.16982	*****
14	0.02079	*****
15	-0.01251	*****
16	-0.02433	*****
17	-0.11583	*****
18	0.00575	*****
19	0.06182	*****
20	0.00673	*****
21	-0.09519	*****
22	0.00208	*****
23	0.00733	*****
24	0.01842	*****

<표 2.2>  $\nabla Y_t$ 의 SACF와 SPACF

Name of Variable = yt		
Period(s) of Differencing		1
Mean of Working Series		40.225
Standard Deviation		264.2074
Number of Observations		80
Observation(s) eliminated by differencing		1

Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 Std Error
0	1.00000	***** 0
1	-0.39953	***** 0.111803
2	-0.16659	***** 0.128416
3	-0.36749	***** 0.131089
4	0.91399	***** 0.143390
5	-0.40224	***** 0.203580
6	-0.14986	***** 0.213284
7	-0.34062	***** 0.214596
8	0.86266	***** 0.221251
9	-0.37132	***** 0.259916
10	-0.14327	***** 0.266465
11	-0.31975	***** 0.267426
12	0.80956	***** 0.272163
13	-0.35702	***** 0.300761
14	-0.12184	***** 0.306013
15	-0.30046	***** 0.306619
16	0.76351	***** 0.310277
17	-0.34027	***** 0.332935
18	-0.11371	***** 0.337254
19	-0.30174	***** 0.337733
20	0.72543	***** 0.341086
21	-0.32671	***** 0.359855
22	-0.10092	***** 0.363544
23	-0.27725	***** 0.363894
24	0.69302	***** 0.366525

Partial Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.39953	*****
2	-0.38818	*****
3	-0.86161	*****
4	0.53703	*****
5	-0.07124	*****
6	-0.07852	*****
7	-0.18205	*****
8	-0.00894	*****
9	0.13990	*****
10	0.00620	*****
11	-0.01644	*****
12	-0.02779	*****
13	-0.05444	*****
14	0.10413	*****
15	0.03459	*****
16	0.06323	*****
17	0.04622	*****
18	-0.01353	*****
19	-0.09883	*****
20	-0.03795	*****
21	-0.06939	*****
22	-0.05790	*****
23	0.03325	*****
24	0.01136	*****

<표 2.2>는  $(1-B)Y_t$ 에 대한 SACF와 SPACF이다. SACF의 주기가 4의 배수일 때 유의한 값이 매우 느리게 감소하고 있는 것으로 보아 차수 4의 계절 차분이 필요한 것으로 보인다.

<표 2.3>  $\nabla_4 \nabla Y_t$ 의 SACF와 SPACF

Name of Variable = yt	
Period(s) of Differencing	1,4
Mean of Working Series	-1.25
Standard Deviation	77.88877
Number of Observations	76
Observation(s) eliminated by differencing	5

Autocorrelations											
Lag	Correlation	1	2	3	4	5	6	7	8	9	Std Error
0	1.00000										0
1	0.07923										0.114708
2	0.01061										0.115426
3	-0.07585										0.115438
4	-0.49349										0.116092
5	-0.19723										0.141019
6	-0.05252										0.144603
7	0.07775										0.144853
8	0.20616										0.145401
9	0.14412										0.149198
10	-0.09251										0.151019
11	0.02835										0.151762
12	-0.16504										0.151832
13	-0.08280										0.154175
14	0.11135										0.154759
15	0.05856										0.155809
16	0.13425										0.156099
17	0.01500										0.157610
18	-0.02802										0.157629
19	-0.07577										0.157695
20	-0.28811										0.158173
21	-0.05064										0.164934
22	0.07393										0.165138
23	0.01801										0.165573
24	0.21307										0.165599

Partial Autocorrelations											
Lag	Correlation	1	2	3	4	5	6	7	8	9	Std Error
1	0.07923										0.114708
2	0.00436										0.115426
3	-0.07752										0.115438
4	-0.48743										0.116092
5	-0.18135										0.141019
6	-0.05608										0.144603
7	0.03078										0.144853
8	-0.06424										0.145401
9	-0.05050										0.149198
10	-0.23383										0.151019
11	0.07051										0.151762
12	-0.10629										0.151832
13	-0.02268										0.154175
14	-0.01336										0.154759
15	0.07444										0.155809
16	0.02605										0.156099
17	-0.03434										0.157610
18	0.02972										0.157629
19	0.04089										0.157695
20	-0.28793										0.158173
21	-0.03389										0.164934
22	0.04291										0.165138
23	-0.05124										0.165573
24	-0.06206										0.165599

<표 2.3>을 보면 1차 차분과 계절 차분의 결과  $\nabla_4 \nabla Y_t$ 는 Seasonal AR(1) 모형을 따른다고 볼 수 있다. 따라서 식별한 모형은  $ARIMA(0,1,0) \times (1,1,0)_4$ 이다.

### 2.1.3 모형 추정

추세모수  $\theta_0$ 는  $|-1.25/77.88877 * \sqrt{76}| = 0.1399 \leq 2$ 이므로 모형에 포함시키지 않는다. noconstant로 모형을 추정해보면 <표 2.4>과 같다.

<표 2.4> 최우추정

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Pr >  t	Lag
AR1,1	-0.50665	0.09772	-5.18	<.0001	4
Variance Estimate		4530.1			
Std Error Estimate		67.30602			
AIC		857.6645			
SBC		859.9953			
Number of Residuals		76			

최우추정 방법으로 추정한 결과  $\hat{\phi} = -0.50665$ 로 유의하게 나왔다.

#### 2.1.4 모형 진단

잔차들에 대한 유의성 검정 결과 <표 2.5>에 주어진  $Pr > ChiSq$ 값이 0.4488, 0.5731, 0.7702, 0.4062로 모두 오차항에 대한 가정을 만족하고 있다고 볼 수 있으며, <표 2.6>에 있는 추정된 잔차들의 SACF와 SPACF도 패턴을 보이고 있지 않아 백색잡음과정을 따른다고 할 수 있다. 시차 20이 조금 유의한 값이 나오는 듯 하지만 이 정도는 무시하고 넘어가도 무방하다고 판단하였다. 그러므로 최종 모형은 식 (2.1)과 같다.

$$(1 + 0.50665B^4)(1 - B^4)(1 - B)Y_t = a_t \quad (\text{식 2.1})$$

<표 2.5> AR(1) 모형 적합에 대한 포트맨토우 검정결과

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.74	5	0.4488	-0.044	0.002	-0.036	-0.018	-0.178	-0.146
12	9.53	11	0.5731	0.095	-0.098	0.049	-0.111	0.135	-0.048
18	12.48	17	0.7702	-0.022	0.098	0.073	-0.073	-0.065	0.070
24	23.96	23	0.4062	-0.064	-0.249	-0.032	0.165	-0.069	0.087

<표 2.6> 잔차의 SACF와 SPACF

Autocorrelation Plot of Residuals														
Lag	Correlation	1	2	3	4	5	6	7	8	9	10	11	12	13
0	1.00000													
1	-0.04429		*											
2	0.00224													
3	-0.03599		*											
4	-0.01775													
5	-0.17802		****											
6	-0.14597		***											
7	0.09469			**										
8	-0.09846			**										
9	0.04884			*										
10	-0.11083			**										
11	0.13460				***									
12	-0.04773			*										
13	-0.02231													
14	0.09841				**									
15	0.07309				*									
16	-0.07316			*										
17	-0.06499			*										
18	0.07036			*										
19	-0.06371			*										
20	-0.24891			****										
21	-0.03155			*										
22	0.16526				***									
23	-0.06912			*										
24	0.08665			*										

Partial Autocorrelations														
Lag	Correlation	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-0.04429													
2	0.00028													
3	-0.03595			*										
4	-0.02100													
5	-0.18038				***									
6	-0.17021				*									
7	0.07670					**								
8	-0.11041					*								
9	0.01735													
10	-0.15418					*								
11	0.06545					*								
12	-0.04318					*								
13	-0.05016					*								
14	0.07497					*								
15	0.06559					*								
16	-0.09461					*								
17	-0.02632					*								
18	0.01039					*								
19	-0.00403					*								
20	-0.27088					****								
21	-0.06725					*								
22	0.12850					*								
23	-0.08310					*								
24	0.06757					*								

인구수를 예측한 결과는 <표 2.7>과 같다.

<표 2.7> 모형 예측 결과

Forecasts for variable yt				
Obs	Forecast	Std Error	95% Confidence Limits	
82	12107.5864	67.3060	11975.6690	12239.5038
83	11945.2143	95.1851	11758.6549	12131.7736
84	12369.0016	116.5774	12140.5140	12597.4892
85	12360.6013	134.6120	12096.7666	12624.4360

## 2.2 분해법

분해법의 관점에서 시도표를 살펴보면, 계절변동이 규칙적이며 선형추세에 의존하지 않고 진폭이 일정한 형태를 띠고 있어 가법모형이 적절하다고 생각된다. 즉  $Y_t = T_t + S_t + I_t$ 를 따른다고 하겠다. 앞서 ARIMA모형을 추정한 결과 자기상관이 뚜렷하므로 곧바로 자기회귀오차모형에 적합시켰고 그 결과는 <표 2.8>과 같다.

<표 2.8> 자기회귀오차모형의 추정 결과

Maximum Likelihood Estimates					
SSE	291250.524	DFE			73
MSE	3990	Root MSE			63.16433
SBC	930.252115	AIC			911.096521
MAE	44.8329586	AICC			913.096521
MAPE	0.4402855	HQC			918.781996
Log Likelihood	-447.54826	Transformed Regression R-Square			0.9191
Durbin-Watson	1.9242	Total R-Square			0.9957
Pr < DW	0.3657	Observations			81
Pr > DW	0.6343				

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	9004	105.6041	85.26	<.0001
t	1	39.0475	2.1609	18.07	<.0001
ID1	1	-68.7661	20.0672	-3.43	0.0010
ID2	1	-163.9403	23.1753	-7.07	<.0001
ID3	1	-423.5029	20.0954	-21.07	<.0001
AR1	1	-0.8277	0.0778	-10.64	<.0001
AR4	1	-0.3524	0.1196	-2.95	0.0043
AR5	1	0.2934	0.1140	2.57	0.0121

NLAG=8로 오차에 AR(8) 모형을 적합한 후 BACKSTEP 옵션으로 유의하지 않은 차수 2,3,6,7,8이 제거되고 최종적으로 시차 1,4,5만이 유의한 것으로 판정되었다. 추정된 값들은 모두 유의한 결과를 보이고 있고, DW검정의 유의확률은 0.3657로 더 이상 자기상관이 존재하지 않는다고 판단할 수 있게 되었다. 최종으로 추정된 자기회귀오차를 가진 회귀모형은 (식 2.2)와 같다,

$$\hat{Y}_t = 9004 + 39.0475t - 68.7661ID_{t1} - 163.9403ID_{t2} - 423.5029ID_{t3} + \hat{e}_t$$

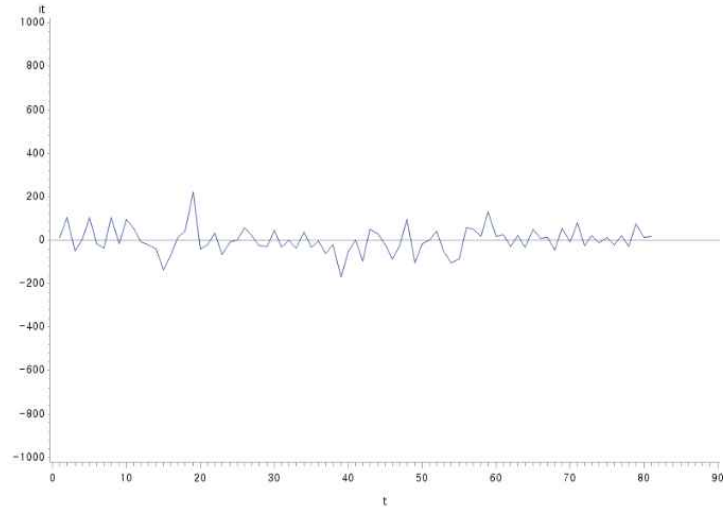
$$\hat{e}_t = 0.8277\hat{e}_{t-1} + 0.3524\hat{e}_{t-4} - 0.2934\hat{e}_{t-5} \quad (\text{식 2.2})$$

추정된 추세·계절성분과 불규칙성분은 <표 2.9>에 주어져있고, 불규칙성분의 시도표를 그려보면 <그림 2.2>에 보이는 것처럼 0을 중심으로 랜덤한 패턴을 보이고 있으므로 성분추정이 잘 되었음을 나타내고 있다.

<표 2.9> 각 성분의 추정 결과

t	yt	ttst	it	fore
1	8988	8974.45	13.548	8974.28
2	9033	8930.26	102.744	8918.15
3	8742	8791.95	-49.947	8697.64
4	9200	9197.20	2.799	9160.19
5	9268	9163.68	104.317	9130.47
(중략)				
76	12022	12042.26	-20.259	11971.61
77	12021	12000.32	20.680	11941.89
78	11916	11945.14	-29.136	11885.76
79	11782	11705.45	76.554	11665.25
80	12222	12211.43	10.570	12127.80
81	12206	12189.17	16.830	12098.08

<그림 2.2> 불규칙성분의 시도표

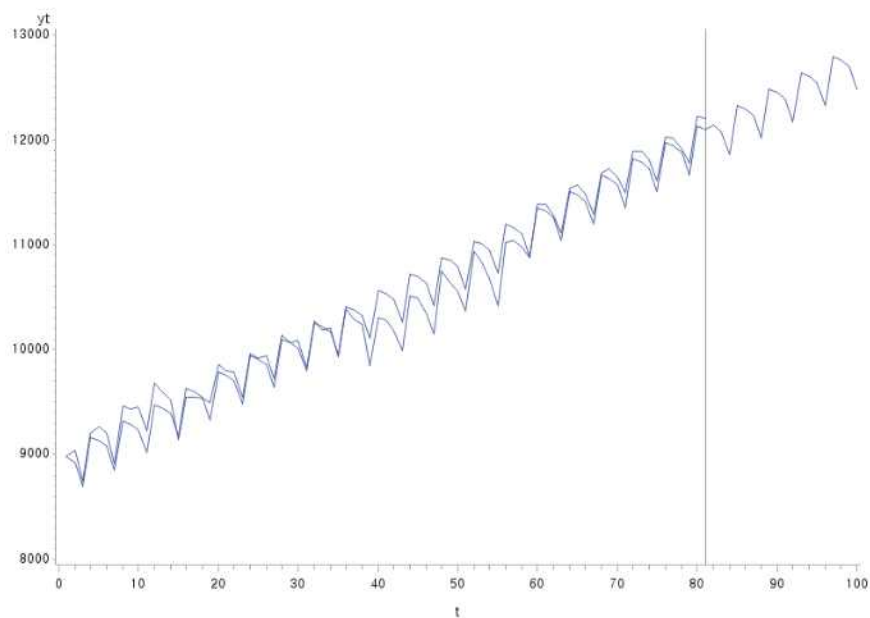


분해법에 의한 2019년 4/4분기부터 2024년 2/4분기까지의 예측값이 <표 2.10>에 나와 있고, 원계열과 예측계열의 시도표를 <그림 2.3>에서 살펴보면 예측이 잘 되고 있음을 알 수 있다.

<표 2.10> 2019년 4/4분기~2024년 2/4분기까지 예측값

t	fore
82	12137.13
83	12081.00
84	11860.49
85	12323.04
86	12293.32
87	12237.19
88	12016.68
89	12479.23
90	12449.51
91	12393.38
92	12172.87
93	12635.42
94	12605.70
95	12549.57
96	12329.06
97	12791.61
98	12761.89
99	12705.76
100	12485.25

<그림 2.3> 2019년 4/4분기~2024년 2/4분기까지의 예측 시도표



### 3. 분석 결과

앞서 분석한 두 가지 모형을 비교하고자 AIC 통계량을 사용하겠다. ARIMA 모형으로 분석하였을 때의 AIC는 857.6645, 분해법으로 분석하였을 때의 AIC는 911.096521로 이 자료에는 분해법보다는 ARIMA 모형으로 분석하는 것이 더 적합하다고 볼 수 있다.

### 4. 결론

우리나라 여성의 분기별 경제 활동 인구수를 위와 같은 절차로 분석하여 예측해본 결과, ARIMA모형과 분해법 모두 적합한 모형을 찾을 수 있었다. 이 분석의 결과로 우리나라 여성 경제 활동 인구수는 분기별 계절성을 띠고 있으며, 점차 증가하는 추세임을 알 수 있다. 비록 경력단절여성과 같은 문제는 아직 해결 과정 중에 있지만, 경제 활동에 참여하는 여성이 늘고 있고 앞으로도 늘어날 것이라는 인사이트는 여성 인권이 더 이상 하위 계층에 머물러 있지는 않을 것이라는 시사점에 도달하도록 이끌어준다.

## 부록

```
data women;
infile 'C:\Users\Sungshin\Desktop\여성경제활동인구.txt';
input yt;
t=_n_;
run;
proc gplot data=women;
symbol1 i=join;
plot yt*t;
run;
proc arima data=women;
identify var=yt nlag=24;
identify var=yt(1) nlag=24;
identify var=yt(1,4) nlag=24;
estimate p=(4) ml plot noconstant;
forecast lead=4;
run;
```

```
data women1; /*t, yt, ID1, ID2, ID3*/
set women;
ID1=mod(t,4)=1;
ID2=mod(t,4)=2;
ID3=mod(t,4)=3;
run;
proc autoreg data=women1;
model yt=t ID1 ID2 ID3 / dwprob nlag=8 method=ML
backstep;
```



```

output out=out1 p=ttst;
run;
data women2; /*t, yt, ttst, it, ID1, ID2, ID3*/
set out1;
it=yt-ttst;
run;
proc gplot data=women2;
axis1 order=(-1000 to 1000 by 200);
plot it*t=1/vref=0 vaxis=axis1 noframe;
run;

data forecast1;
set women2;
t=_n_+81;
if t>100 then delete;
keep ID1 ID2 ID3 t;
run;
proc print data=forecast1;
run;

data forecast2;
set women2 forecast1;
fore=9004+39.0475*t-68.7661*ID1-163.9403*ID2-423.5029
*ID3;
run;
proc print data=forecast2;
run;

```

```
proc gplot data=forecast2;  
symbol1 i=join line=1;  
symbol2 i=join line=2;  
plot yt*t=1 fore*t=1/overlay noframe href=81;  
run;
```