8장 Advanced ARIMA 모델

Factored ARIMA model

- A factored model (also referred to as a multiplicative model) represents the ARIMA model as a product of simpler ARIMA models.
- Popularly applied to modeling seasonality
- **For example**
  - you might model *SALES* as a combination of an AR(1) process that reflects short term dependencies and an AR(12) model that reflects the seasonal pattern.
  - It might seem that the way to do this is with the option P=(1 12), but the AR(1) process also operates in past years; you really need autoregressive parameters at lags 1, 12, and 13.
  - You can specify a subset model with separate parameters at these lags, or you can specify a factored model that represents the model as the product of an AR(1) model and an AR(12) model.
  - Consider the following two ESTIMATE statements in SAS.

Factored ARIMA model

```
proc arima data=call;
identify var=sales ;
estimate method=ML
p=(1 12 13);
run;
```

```
proc arima data=call;
identify var=sales ;
estimate method=ML
p=(1)( 12);
run;
```

- The mathematical form of the autoregressive models produced by these two specifications are shown as

Option	Autoregressive Operator
P=(1 12 13)	$(1 - \phi_1 B - \phi_{12} B^{12} - \phi_{13} B^{13})$
P=(1)(12)	$(1 - \phi_1 B)(1 - \phi_{12} B^{12})$

Seasonal ARIMA model

- The multiplicative double seasonal ARIMA model can be written as

$$\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})(1-B)^d(y_t - c) = \theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})\varepsilon_t,$$

- where c is a constant, and ε_t is a white noise error term.
- B is the back shift operator with $B^l y_t = y_{t-l}$.
- $\phi_p(B)$ and $\theta_q(B)$ indicate autoregressive (AR) and moving average (MA) part
- They are polynomial functions of orders p and q, respectively.
- $\Phi_{P_1}(B^{S_1})$ and $\Phi_{P_2}(B^{S_2})$ indicate two seasonal factors related with the AR part.
- $\Theta_{Q_1}(B^{S_1})$ and $\Theta_{Q_2}(B^{S_2})$ indicate two seasonal factors related with the MA part.
- P_1, P_2, Q_1, and Q_2 indicate the order of corresponding polynomial functions.

Seasonal ARIMA model

- s_1 and s_2 indicate the length of two seasonal cycles.
- Note that the ARIMA model is reduced to the autoregressive moving average (ARMA) model if d =0 (no differencing). A model fitted with differencing (d = 1) would be useful if there is evidence of significant trend effects.
- **For example**
  - Forecasting intraday 5-minute call arrivals at a call center.
  - We consider the double seasonality such as intraday and intraweek cycles.
  - Therefore, s_1 =169 and s_2 =845 were used.
  - Their corresponding periods indicate the lengths of a day and a week.
  - For the selection of lag order in the model, all parameter estimates were requested to be significant at level 1% using the maximum likelihood estimator (MLE).
  - Insignificant lags were removed from the model.
  - Additionally, the Schwarz Bayesian criterion (SBC) was applied to the model selection.

Seasonal ARIMA model

- Since the evidence of trend effect was not detected in the unit-root tests, a model with no differencing was considered.

- The final selected double seasonal ARMA model based on the in-sample real data is written as

$$(1-\phi_1 B-\phi_2 B^2)(1-\phi_{169}B^{169})(1-\phi_{845}B^{845})(y_t-c)$$
$$=(1-\theta_1 B)(1-\theta_{169}B^{169})(1-\theta_{845}B^{845})\varepsilon_t \,,$$

- estimation results are

Parameter estimates in Models 4 for the real data

Lag, *i*		1	2	169	845
(Model 4)					
ϕ_i		1.0192**	-0.0329*	0.9975**	0.9972**
θ_i		0.7889**		0.9682**	0.9836**
c	13.4838**				

Note: **: P<0.0001, *: p<0.0025

Seasonal ARIMA model

– SAS codes are

```
proc arima data=call;
identify var=arrival ;
estimate method=ML
p=(1 2) (169 )(845)
q=(1) (169 ) (845) ;
forecast lead=169 interval=1 id=seq out=cast alpha=0.05;
run;
quit;
```

ARMAX model

Auto Regressive Moving Average with Exogeneous Variable (ARMAX) model

- If y_t depends on an exogeneous variable x_t in some way, a dynamic regression model may be written as

$$y_t = c + f(x_t) + e_t,$$

where c is a constant term.

- A mathematical function $f(\cdot)$ is referred to as a transfer function.
- e_t is an auto-correlated stochastic error term, which is independent of x_t.
- For example, e_t are assumed as

$$e_t = \frac{\theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})}{\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})}\varepsilon_t, \quad (1)$$

where ε_t is a white noise.

- If ε_t is assumed to follow a normal distribution, then the maximum likelihood (ML) method for the parameter estimation of the models would be available.

ARMAX model

- Equation (1) indicates that the error term e_t is modeled via a double SARMA mechanism.
- B is the back shift operator with $B^l y_t = y_{t-l}$.
- $\phi_p(B)$ and $\theta_q(B)$ indicate the autoregressive (AR) and moving average (MA) parts, respectively. They are polynomial functions of orders p and q, respectively.
- $\Phi_{P_1}(B^{S_1})$ and $\Phi_{P_2}(B^{S_2})$ indicate two seasonal factors related to the AR part.
- $\Theta_{Q_1}(B^{S_1})$ and $\Theta_{Q_2}(B^{S_2})$ indicate two seasonal factors related to the MA part.
- P_1, P_2, Q_1, and Q_2 indicate the order of corresponding polynomial functions
- s_1 and s_2 indicate the length of two seasonal cycles.
- This stochastic disturbance is often represented as $\text{ARMA}(p,q)\times(P_1,Q_1)_{s_1}\times(P_2,Q_2)_{s_2}$.
- The dynamic regression models with SARMA errors are often referred to as SARMA with the exogeneous variable (SARMAX) model.

ARMAX model

- **For example**
  - Forecasting intraday hourly electricity demand using The SARMAX model with holiday effects
  - dummy variables are employed to identify holidays.
  - we applied daily based special day effects to the dynamic regression model with the SARMA error term.

$$y_t = c + \gamma_0 d_t + \frac{\theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})}{\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})}\varepsilon_t ,$$

where d_t is an exogeneous variable representing the special day effects on a daily basis in the SARMAX model.

  - $d_t = 1$ if time t belongs to holidays
  - Here, all the hours of special days are coded into 1 because the special day effects are coded on a daily basis. At all the other time, d_t is coded into 0.

ARMAX model

- The ML method is applied to the parameter estimation of all the models under the assumption of normality of error term. All ML estimates were requested to be significant at level 1%.
- Insignificant variables are deleted from the model.
- Furthermore, the Akaike information criterion (AIC) and Schwarz Bayesian criterion (SBC) are implemented for the model selection.
- Parameter estimation results are reported as

Lag	1	2	5	12	24	48	72	96	144	168	672	holiday
AR	1.495	-0.543	0.016	0.012	0.325		0.111	0.114	0.396	1.248	-0.248	-0.308
MA	0.092		-0.065			-0.145			0.329	1.172	-0.178	

ARMAX model

– SAS codes are

```
proc arima data=call;
identify var=demand
crosscorr=(holiday );
estimate method=ML p=(1   2   5   12) ( 24   28   72   96   144) (168 672 )
                   q=(1   5) (48) (168   672)
input=(holiday);
forecast lead=24 interval=1 id=seq out=cast alpha=0.05;
run;
quit;
```

Trend Stationary Model with AR error

- **For Example**
  - You might model *Real GNP* as a linear deterministic time trend with an error following a combination of an AR(1) process that reflects short term dependencies and an AR(10) model that reflects the seasonal pattern.
  - Model can be written down as

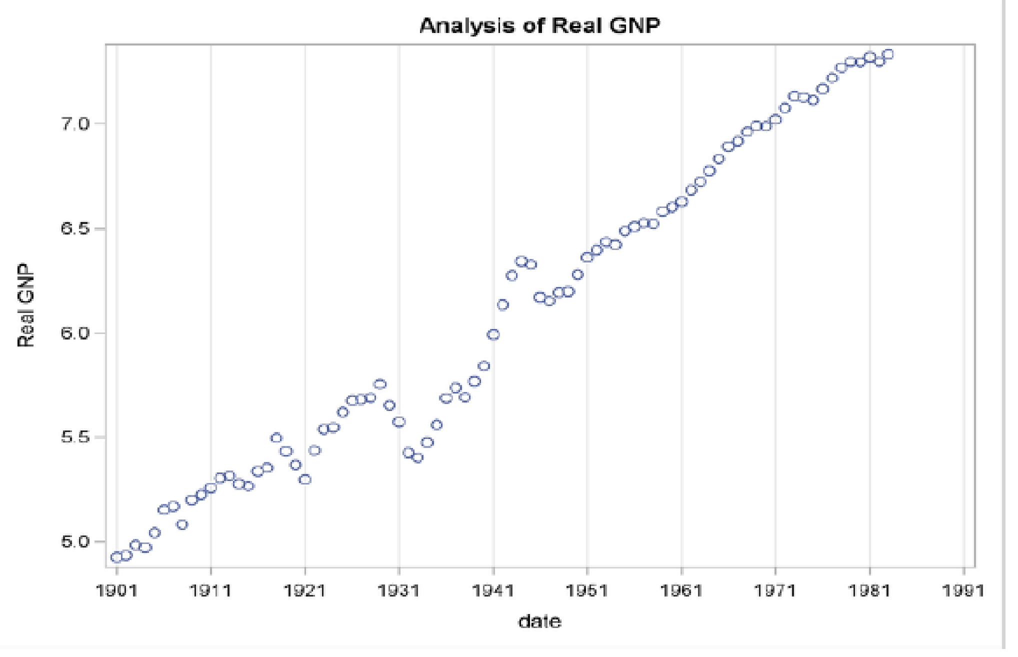
$$y_t = a + bt + u_t$$
$$u_t = \phi_1 u_{t-1} + \phi_4 u_{t-4} + \phi_5 u_{t-5} + \varepsilon_t$$
$$\varepsilon_t \sim IID\, N(0, \sigma^2)$$

  - SAS codes are

```
proc autoreg data=a;
    model gnp = time / nlag=(1)( 4);      /*same as nlag=(1 4 5)*/
run;
```

Trend Stationary Model with AR error

– Figure for GNP is given



– Extension to GARCH model (see Ch. 5)

```
proc autoreg data=a;
        model y = time / nlag=2 garch=(p=1,q=1,type=exp);
run;
```

Reading lists

[1] Cryer, J.D., Chan, K. (2008) Time Series Analysis with Applications in R, Springer, New York, USA.

[2] Kim, M.S. (2011). A comparison of seasonal linear models and seasonal ARIMA models for forecasting intra-day call arrivals, *The Korean Communications in Statistics*, 18, 237-244.