Time Series Foreacasting

윤재경 정재원 전혜민
윤빈나 권지혜 최성웅

KU-BIG

INDEX

1. What we learned
- Tests for Stationarity
- ACF / PACF
- Parameter selection

2. Application

KU-BIG

Tests for Stationarity(AR process) : DF & ADF Test

1. Dickey-Fuller Test

$$y_t = \rho y_{t-1} + u_t$$
$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + u_t$$

$$H_0 : \ \rho = 1(\delta = 0) \ vs \ H_1 : stationarity \ or \ trend - stationarity$$

- Null hypothesis : **Unit root** is present ⇔ The process is **non-stationarity**.

Tests for Stationarity(AR process) : DF & ADF Test

2. Augmented Dickey-Fuller Test

$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + \theta_1 \Delta y_{t-1} + \cdots \theta_{p-1} \Delta y_{t-p+1} + u_t$$

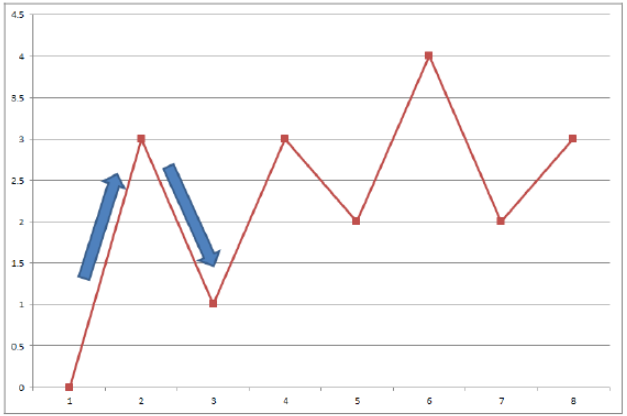
$$H_0 : \delta = 0 \; vs \; H_1 : \delta < 0$$

- Testing higher order of Auto-Regressive processes

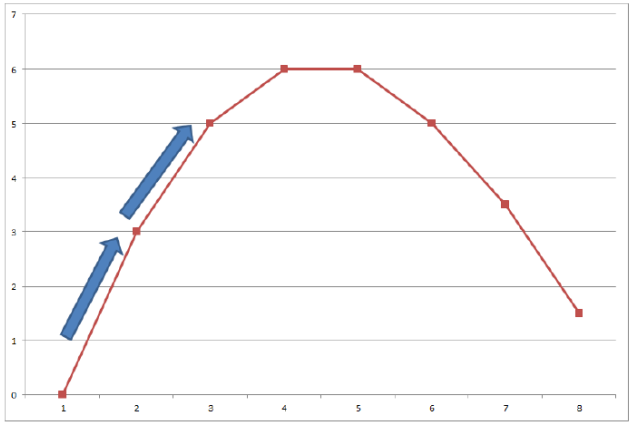
- Null hypothesis : **Unit root** is present ⇔ The process is **non-stationarity**

Autocorrelation Function(ACF)

- ACF : the correlation of a signal with a delayed copy of itself as a function of delay
  - ✓ the similarity between observations as a function of the time lag between them

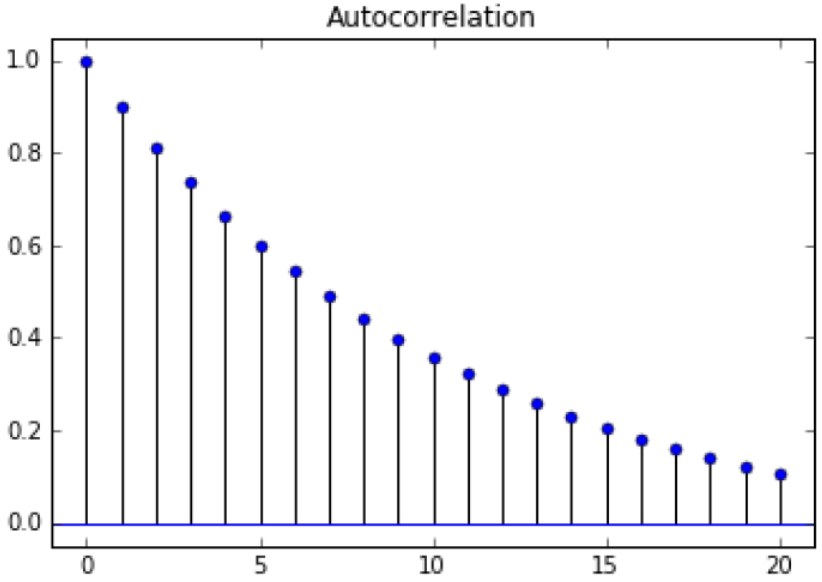


Mean reversion
(negative autocorrelation)

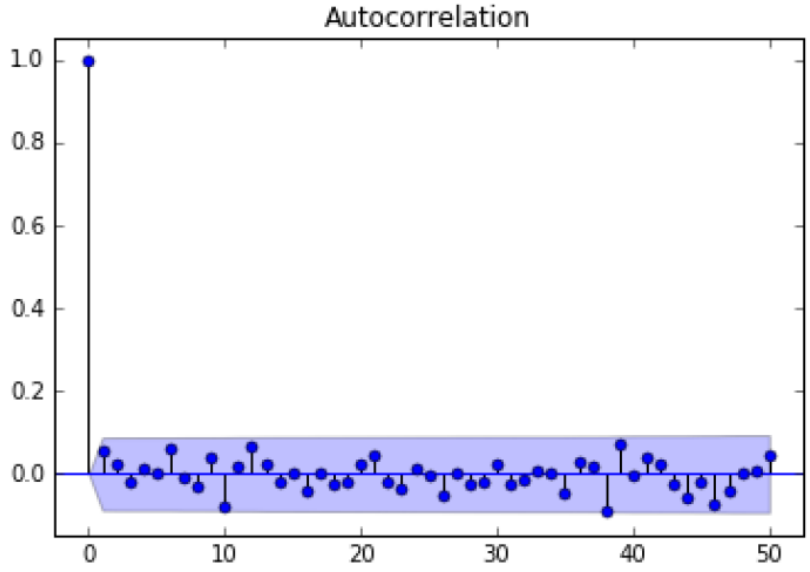


Momentum
(positive autocorrelation)

Autocorrelation Function(ACF)



Momentum
(positive autocorrelation)



White noise
(not significant)

Partial Autocorrelation Function(PACF)

- **partial autocorrelation function (PACF)** gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags
- **partial correlation :** the degree of association between two random variables, with the effect of a set of controlling random variables removed

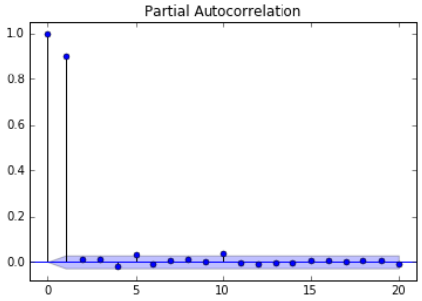
$$R_t = \phi_{0,3} + \phi_{1,3} R_{t-1} + \phi_{2,3} R_{t-2} + \phi_{3,3} R_{t-3} + \epsilon_{3t}$$

$$R_t = \phi_{0,4} + \phi_{1,4} R_{t-1} + \phi_{2,4} R_{t-2} + \phi_{3,4} R_{t-3} + \phi_{4,4} R_{t-4} + \epsilon_{4t}$$

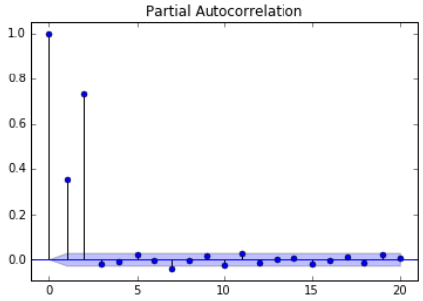
$$\vdots$$

Partial Autocorrelation Function(PACF) - parameter selection

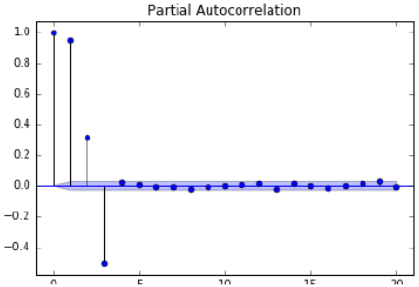
- AR(1)



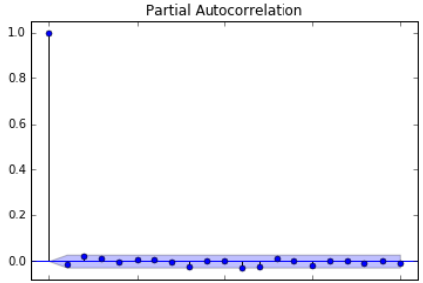
- AR(2)



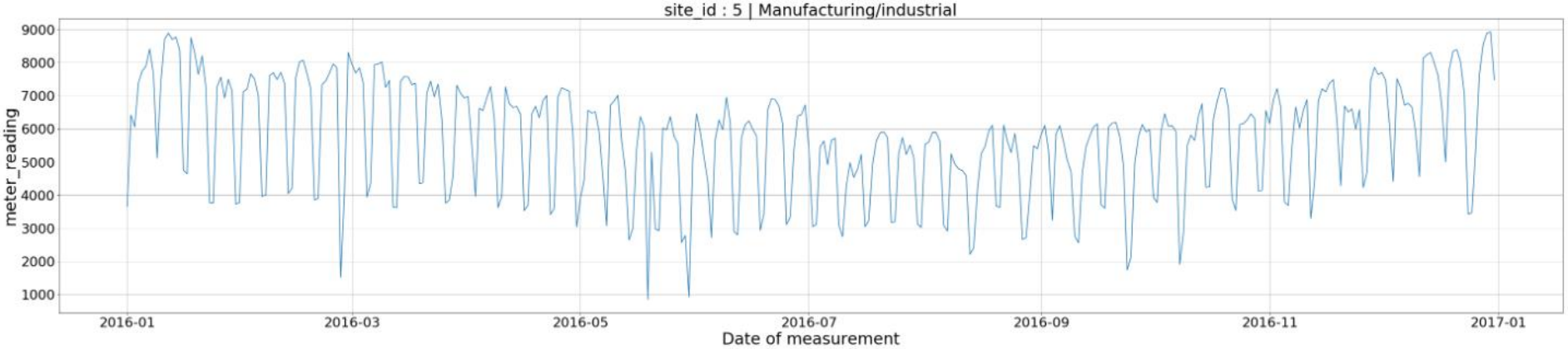
- AR(3)



- White Noise



2. Application



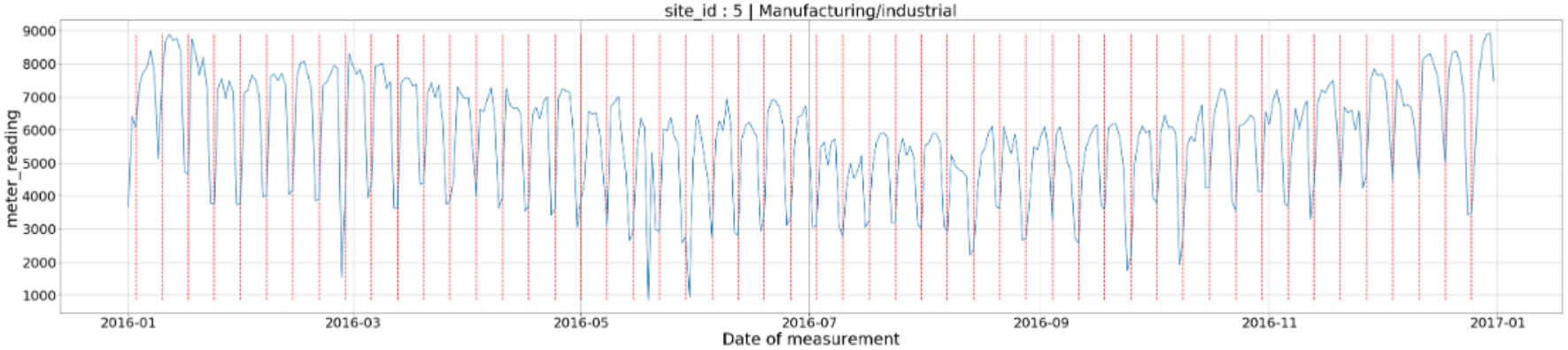
2. Application

In [17]:

```
result = adfuller(tmp,regression='nc', autolag='t-stat')
print('regression :ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
regression :ADF Statistic: -0.430076
p-value: 0.524622
```

2. Application



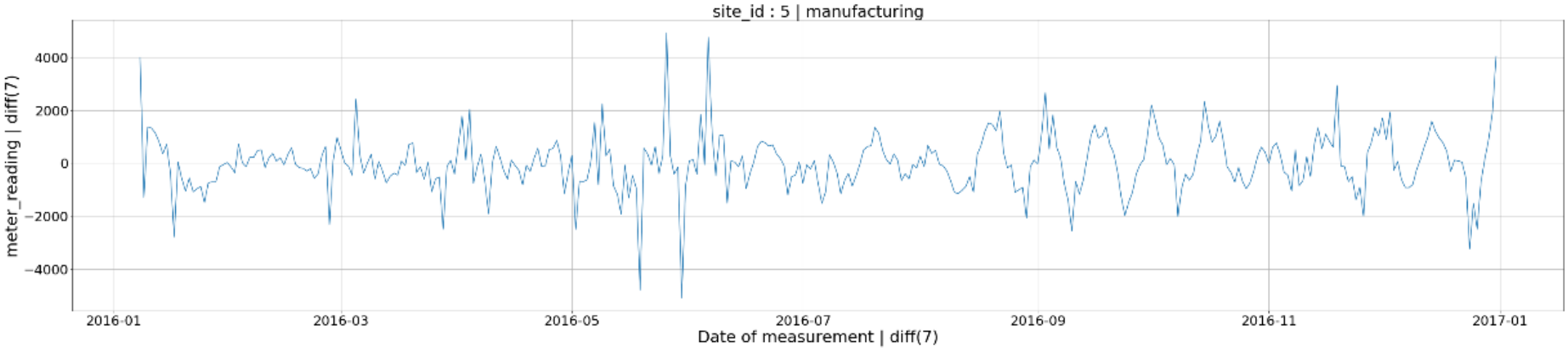
2. Application

In [21]:
```
tmp_sa = tmp.diff(7)
```

In [24]:
```
result = adfuller(tmp_sa,regression='nc', autolag='t-stat')
print('regression :ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

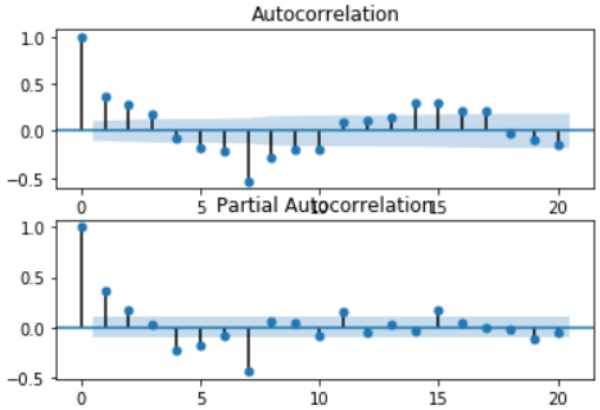
```
regression :ADF Statistic: -4.075720
p-value: 0.000054
```

2. Application



In [20]:

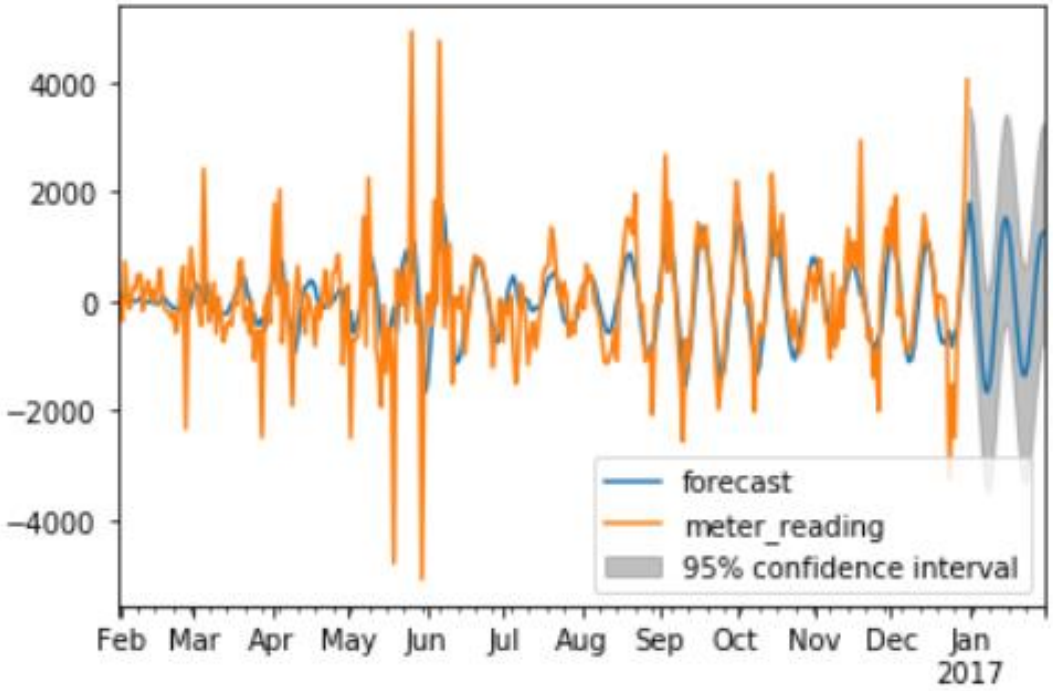
```
#acf, pacf of seasonality-removed graph
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
fig, axes = plt.subplots(2,1)
plot_acf(tmp_sa, lags=20, ax=axes[0])
plot_pacf(tmp_sa, lags=20, ax=axes[1])
plt.show()
```



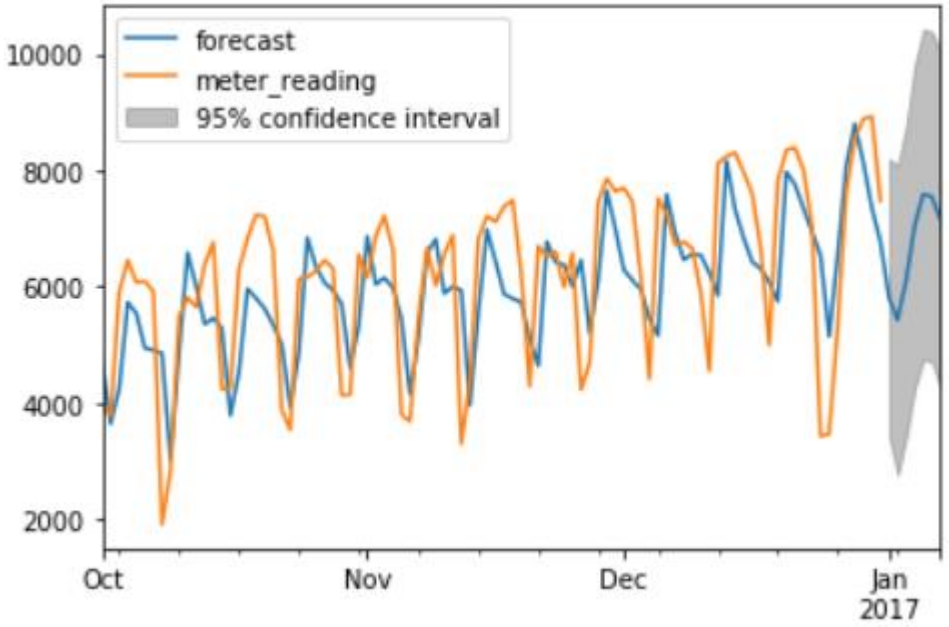
▲ AR차수를 2로 설정

```
the AIC for AR(1):  5986.35461222374
the AIC for AR(2):  5978.3898404295805
the AIC for ARMA(2,1):  5980.295578351251
the AIC for ARMA(2,2):  5913.903623062178
the AIC for ARMA(2,3):  5933.580468658558
the AIC for ARMA(2,4):  5913.743469610157
the AIC for ARMA(2,5):  5915.521129662845
the AIC for ARMA(2,6):  5916.937497841946
```

▲ARMA(2,2) 으로 model select


4000
2000
0
−2000
−4000
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2017
forecast
meter_reading
95% confidence interval

```
the AIC for ARIMA(2,1,0): 6383.458844406732
the AIC for ARIMA(2,1,1): 6264.48947185209
the AIC for ARIMA(2,1,2): 6237.960076322313
the AIC for ARIMA(2,2,0): 6543.2890185163615
the AIC for ARIMA(2,2,2): 6248.480230306044
```


forecast
meter_reading
95% confidence interval
10000
8000
6000
4000
2000
Oct
Nov
Dec
Jan
2017


Do you
have any
question?
Thank you
for your attention.