In [1]:
```
import pandas as pd
from sklearn import linear_model
import matplotlib.pyplot as plt
import numpy as np
```

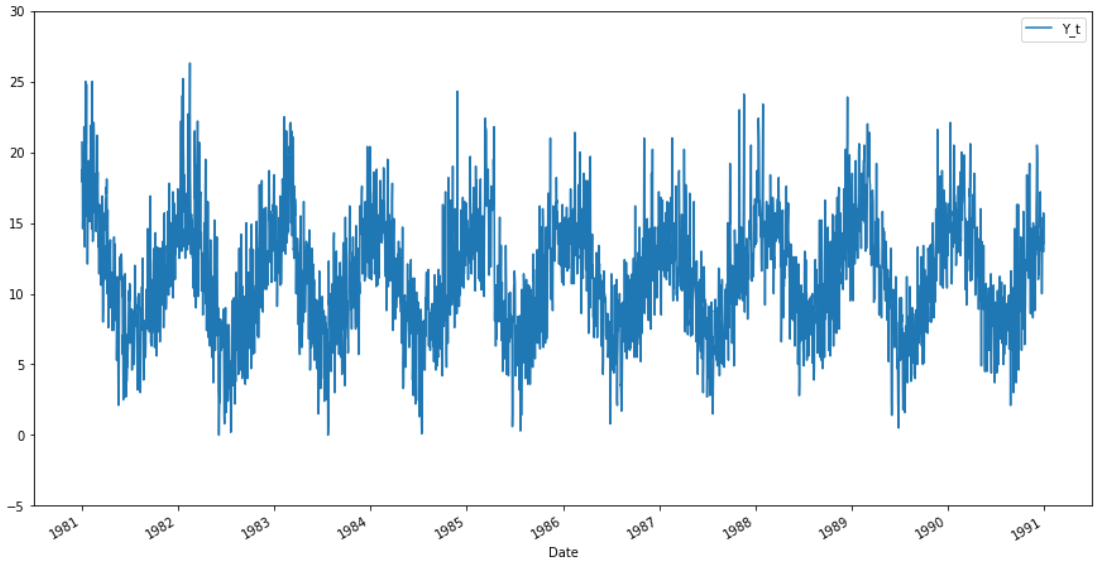
Ελάχιστες ημερήσιες θερμοκρασίες που παρατηρήθηκαν στη Μελβούρνη από 01-01-1981 μέχρι 31-12-1990

In [2]:
```
dataset = pd.read_csv('min_temp.csv', header=0, infer_datetime_format=True, par
se_dates=[0], index_col=[0])
```

In [3]:
```
dataset = dataset.rename(columns={'Temp':'Y_t'})
```

In [4]:
```
dataset.plot(figsize=(15,8))
plt.ylim((-5,30))
```

Out[4]: (-5, 30)



In [5]:
```
dataset
```

Out[5]:

Date	Y_t
1981-01-01	20.7
1981-01-02	17.9
1981-01-03	18.8
1981-01-04	14.6
1981-01-05	15.8
...	...
1990-12-27	14.0
1990-12-28	13.6
1990-12-29	13.5
1990-12-30	15.7
1990-12-31	13.0

3650 rows × 1 columns

Χωρισμός του dataset σε training και testing set

training -> 01-01-1981 μέχρι 21-12-1990

testing -> 22-12-1990 μέχρι 31-12-1990

In [6]:
```
dataset['Y_(t-1)'] = dataset['Y_t'].shift(1)
dataset['Y_(t-2)'] = dataset['Y_t'].shift(2)
```

In [7]:
```
dataset
```

Out[7]:

Date	Y_t	Y_(t-1)	Y_(t-2)
1981-01-01	20.7	NaN	NaN
1981-01-02	17.9	20.7	NaN
1981-01-03	18.8	17.9	20.7
1981-01-04	14.6	18.8	17.9
1981-01-05	15.8	14.6	18.8
...	...	...	...
1990-12-27	14.0	14.6	12.9
1990-12-28	13.6	14.0	14.6
1990-12-29	13.5	13.6	14.0
1990-12-30	15.7	13.5	13.6
1990-12-31	13.0	15.7	13.5

3650 rows × 3 columns

In [8]:
```
training_set = pd.DataFrame(dataset[:3640])
testing_set = pd.DataFrame(dataset[3640::])

training_set = training_set.drop(dataset.index[[0,1]])
```

In [9]:
```
training_set
```

Out[9]:

Date	Y_t	Y_(t-1)	Y_(t-2)
1981-01-03	18.8	17.9	20.7
1981-01-04	14.6	18.8	17.9
1981-01-05	15.8	14.6	18.8
1981-01-06	15.8	15.8	14.6
1981-01-07	15.8	15.8	15.8
...	...	...	...
1990-12-17	13.9	13.6	13.4
1990-12-18	17.2	13.9	13.6
1990-12-19	14.7	17.2	13.9
1990-12-20	15.4	14.7	17.2
1990-12-21	13.1	15.4	14.7

3638 rows × 3 columns

In [10]:
```
testing_set
```

Out[10]:

Date	Y_t	Y_(t-1)	Y_(t-2)
1990-12-22	13.2	13.1	15.4
1990-12-23	13.9	13.2	13.1
1990-12-24	10.0	13.9	13.2
1990-12-25	12.9	10.0	13.9
1990-12-26	14.6	12.9	10.0
1990-12-27	14.0	14.6	12.9
1990-12-28	13.6	14.0	14.6
1990-12-29	13.5	13.6	14.0
1990-12-30	15.7	13.5	13.6
1990-12-31	13.0	15.7	13.5

Συντελεστής γραμμικής συσχέτισης για το training set

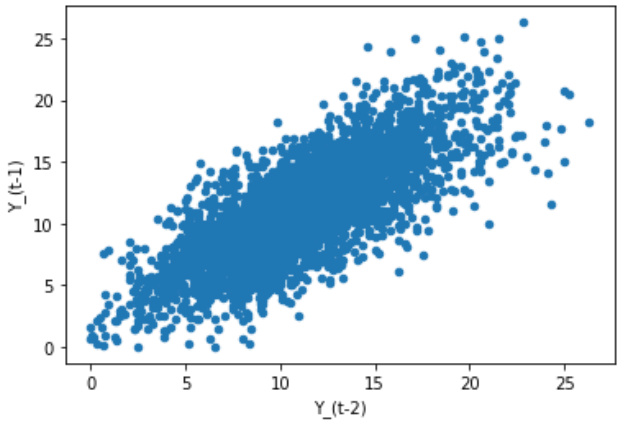
In [11]:
```
training_set.corr()
```

Out[11]:

	Y_t	Y_(t-1)	Y_(t-2)
Y_t	1.000000	0.774745	0.631066
Y_(t-1)	0.774745	1.000000	0.774938
Y_(t-2)	0.631066	0.774938	1.000000

In [15]:
```
training_set.plot.scatter(x='Y_(t-2)', y='Y_(t-1)')
```

Out[15]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7506f16310>
```



Αυτοπαλινδρομικό μοντέλο 2ης τάξης AR(2)

$$Y_t = A + B^{(1)}Y_{t-1} + B^{(2)}Y_{t-2} + \epsilon_t$$

In [16]:
```
model = linear_model.LinearRegression()
X = training_set[['Y_(t-2)','Y_(t-1)']]
y = training_set['Y_t']
model.fit(X,y)
```

Out[16]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [17]:
```
model.coef_
```

Out[17]:
```
array([0.07674685, 0.71496785])
```

In [18]:
```
X = testing_set[['Y_(t-2)','Y_(t-1)']]
testing_set['pr1'] = model.predict(X)
print(testing_set)
```

```
             Y_t  Y_(t-1)  Y_(t-2)        pr1
Date
1990-12-22  13.2     13.1     15.4  12.872760
1990-12-23  13.9     13.2     13.1  12.767739
1990-12-24  10.0     13.9     13.2  13.275892
1990-12-25  12.9     10.0     13.9  10.541240
1990-12-26  14.6     12.9     10.0  12.315334
1990-12-27  14.0     14.6     12.9  13.753345
1990-12-28  13.6     14.0     14.6  13.454834
1990-12-29  13.5     13.6     14.0  13.122799
1990-12-30  15.7     13.5     13.6  13.020603
1990-12-31  13.0     15.7     13.5  14.585858
```

In [19]:
```
Y_tm2 = testing_set['Y_(t-2)'][0]
Y_tm1 = testing_set['Y_(t-1)'][0]

pr2 = []
for i in range(10):
    X = np.array([[Y_tm2, Y_tm1]])
    y_hat = model.predict(X)
    Y_tm2 = Y_tm1
    Y_tm1 = y_hat[0]
    pr2.append(y_hat)

testing_set['pr2']=np.array(pr2)
```

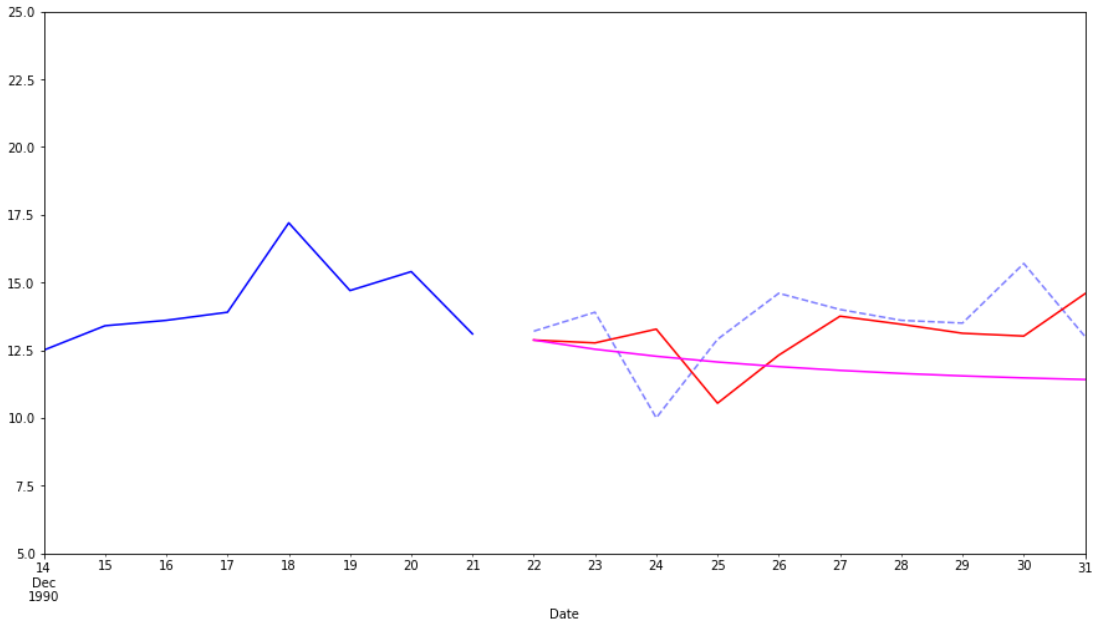
In [20]:
```
testing_set
```

Out[20]:

Date	Y_t	Y_(t-1)	Y_(t-2)	pr1	pr2
1990-12-22	13.2	13.1	15.4	12.872760	12.872760
1990-12-23	13.9	13.2	13.1	12.767739	12.533774
1990-12-24	10.0	13.9	13.2	13.275892	12.273969
1990-12-25	12.9	10.0	13.9	10.541240	12.062201
1990-12-26	14.6	12.9	10.0	12.315334	11.890854
1990-12-27	14.0	14.6	12.9	13.753345	11.752095
1990-12-28	13.6	14.0	14.6	13.454834	11.639736
1990-12-29	13.5	13.6	14.0	13.122799	11.548753
1990-12-30	15.7	13.5	13.6	13.020603	11.475080
1990-12-31	13.0	15.7	13.5	14.585858	11.415424

In [21]:
```
plt.figure(figsize=(15,8))
training_set['Y_t'][3630:].plot(c='blue')
testing_set['Y_t'].plot(c='blue',alpha=0.5, style=['--'])
testing_set['pr1'].plot(c='red')
testing_set['pr2'].plot(c='magenta')
plt.ylim((5,25))
```

Out[21]: (5, 25)



In [29]:
```
model.coef_
```

Out[29]: array([0.77540578])

Υπολογισμός του συντελεστή μερικής αυτοσυσχέτισης PACF(2)

$$Y_t = A_1 + B_1 Y_{t-1} + (\epsilon_1)_t$$
$$\hat{y}_t = a_1 + b_1 y_{t-1}$$

In [22]:
```
X = training_set[['Y_(t-1)']]
y = training_set['Y_t']
model.fit(X,y)
```

Out[22]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

$$(e_1)_t = y_t - \hat{y}_t$$

In [23]:
```
training_set['pr_Y_t|Y_(t-1)'] = model.predict(X)
training_set['e_pr_Y_t|Y_(t-1)'] = training_set['Y_t'] - training_set['pr_Y_t|Y
_(t-1)']
```

In [24]:
```
training_set
```

Out[24]:

Date	Y_t	Y_(t-1)	Y_(t-2)	pr_Y_t|Y_(t-1)	e_pr_Y_t|Y_(t-1)
1981-01-03	18.8	17.9	20.7	16.380553	2.419447
1981-01-04	14.6	18.8	17.9	17.077583	-2.477583
1981-01-05	15.8	14.6	18.8	13.824776	1.975224
1981-01-06	15.8	15.8	14.6	14.754150	1.045850
1981-01-07	15.8	15.8	15.8	14.754150	1.045850
...	...	...	...	...	...
1990-12-17	13.9	13.6	13.4	13.050298	0.849702
1990-12-18	17.2	13.9	13.6	13.282642	3.917358
1990-12-19	14.7	17.2	13.9	15.838418	-1.138418
1990-12-20	15.4	14.7	17.2	13.902224	1.497776
1990-12-21	13.1	15.4	14.7	14.444358	-1.344358

3638 rows × 5 columns

$$Y_{t-2} = A_2 + B_2 Y_{t-1} + (\epsilon_2)_t$$
$$\hat{y}_{t-2} = a_2 + b_2 y_{t-1}$$

In [25]:
```
X = training_set[['Y_(t-1)']]
y = training_set['Y_(t-2)']
model.fit(X,y)
```

Out[25]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

$$(e_2)_{t-2} = y_{t-2} - \hat{y}_{t-2}$$

In [26]:
```
training_set['pr_Y_(t-2)|Y_(t-1)'] = model.predict(X)
training_set['e_pr_Y_(t-2)|Y_(t-1)'] = training_set['Y_(t-2)'] - training_set['
pr_Y_(t-2)|Y_(t-1)']
```

In [27]:
```
training_set
```

Out[27]:

Date	Y_t	Y_(t-1)	Y_(t-2)	pr_Y_t|Y_(t-1)	e_pr_Y_t|Y_(t-1)	pr_Y_(t-2)|Y_(t-1)	e_pr_Y_(t-2)|Y_(t-1)
1981-01-03	18.8	17.9	20.7	16.380553	2.419447	16.389576	4.310424
1981-01-04	14.6	18.8	17.9	17.077583	-2.477583	17.087441	0.812559
1981-01-05	15.8	14.6	18.8	13.824776	1.975224	13.830737	4.969263
1981-01-06	15.8	15.8	14.6	14.754150	1.045850	14.761224	-0.161224
1981-01-07	15.8	15.8	15.8	14.754150	1.045850	14.761224	1.038776
...	...	...	...	...	...	...	...
1990-12-17	13.9	13.6	13.4	13.050298	0.849702	13.055331	0.344669
1990-12-18	17.2	13.9	13.6	13.282642	3.917358	13.287953	0.312047
1990-12-19	14.7	17.2	13.9	15.838418	-1.138418	15.846792	-1.946792
1990-12-20	15.4	14.7	17.2	13.902224	1.497776	13.908277	3.291723
1990-12-21	13.1	15.4	14.7	14.444358	-1.344358	14.451062	0.248938

3638 rows × 7 columns

$$Corr(e_1, e_2)$$

In [28]:
```
training_set[['e_pr_Y_t|Y_(t-1)', 'e_pr_Y_(t-2)|Y_(t-1)']].corr()
```

Out[28]:

	e_pr_Y_t|Y_(t-1)	e_pr_Y_(t-2)|Y_(t-1)
e_pr_Y_t|Y_(t-1)	1.000000	0.076791
e_pr_Y_(t-2)|Y_(t-1)	0.076791	1.000000

In [31]:
```
from statsmodels.tsa.stattools import pacf
print(pacf(dataset['Y_t'][:3640], nlags=10))
```

```
[1.         0.77454064 0.07667323 0.18903945 0.15209657 0.12980496
 0.11015613 0.10214422 0.07452041 0.07049043 0.03490694]
```

$$Y_t = A + B^{(1)}Y_{t-1} + B^{(2)}Y_{t-2} + B^{(3)}Y_{t-3} + \epsilon_t$$