In [1]:

**import** **statsmodels.api** **as** **sm**  
**from** **statsmodels.formula.api** **import** ols  
**import** **pandas** **as** **pd**

In [2]:

root = 'D:/nyu/lag'  
crude\_oil\_df = pd.read\_csv(root + '/crude oil data.csv',index\_col = 'Date', parse\_dates = **True**)  
gasoline\_df = pd.read\_csv(root + '//gasoline weekly data.csv', index\_col = 'Date', parse\_dates = **True**)

In [3]:

codf = crude\_oil\_df.iloc[:,4:5]  
gasdf = gasoline\_df.loc['21-Aug-00':,:]

In [4]:

df = pd.concat([gasdf,codf],axis = 1)  
df.columns = ['gasoline\_price','crude\_oil\_price']  
df.dropna(how = 'any', inplace = **True**)  
df

Out[4]:

|  | **gasoline\_price** | **crude\_oil\_price** |
| --- | --- | --- |
| **Date** |  |  |
| **2000-08-28** | 1.521 | 32.869999 |
| **2000-09-11** | 1.598 | 35.099998 |
| **2000-09-18** | 1.599 | 36.880001 |
| **2000-09-25** | 1.586 | 31.570000 |
| **2000-10-02** | 1.563 | 32.150002 |
| **...** | ... | ... |
| **2020-12-21** | 2.311 | 47.740002 |
| **2020-12-28** | 2.330 | 47.619999 |
| **2021-01-04** | 2.336 | 47.619999 |
| **2021-01-11** | 2.403 | 52.250000 |
| **2021-01-25** | 2.478 | 52.770000 |

960 rows × 2 columns

In [5]:

**from** **statsmodels.graphics.tsaplots** **import** plot\_pacf  
**import** **matplotlib.pyplot** **as** **plt**  
pacf = plot\_pacf(df['gasoline\_price'], lags = 10)  
plt.title('PACF')  
pacf.show()

<ipython-input-5-c52940f00767>:5: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend\_inline, which is a non-GUI backend, so cannot show the figure.  
 pacf.show()

In [6]:

**from** **statsmodels.graphics.tsaplots** **import** plot\_acf  
acf = plot\_acf(df['gasoline\_price'], lags = 10)  
plt.title('ACF')  
pacf.show()

<ipython-input-6-16bf44ad4815>:4: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend\_inline, which is a non-GUI backend, so cannot show the figure.  
 pacf.show()

Explanation:

We use PACF plot and ACF plot with 10 lags to determine the order number of auto-corrrelation of gasoline price.

According to the PACF plot, the first-order and second-order coefficients are high, and the coefficients after the second-order converge to 0 very fast.

According to the ACF plot, within 10 weeks, there is no obvious evidence showing that the coefficients converge to 0.

Hence, we should use the PACF plot to determine the orrder number of auto-correlation of gasoline price and it is AR(2).

This has a real life meaning: In the previous section, we have shown that crude oil price's change is a key predicator of gasoline futures price's change and there is a one-week lag effect of crude oil price's change on gasoline futures price's change. This correspond to the PACF plot that there is a high positive correlationship for the first-order auto-correlation.

However, when it comes the second week, the gasoline futures price has already been above or below the avergae price. If we assume that the gasoline futures market is in a steady state in the long-term, the market participants will notice this and the price will move in the opposite direction of the previous week's moving trend. Hence there is a mean-return effect and we can see that in the PACF plot, there is a high (although a little smaller compared to the first-order) negativce correlationship for the second-order auto-correlation.

In [12]:

CO2\_df = pd.read\_csv(root + '/weekly\_CO2.csv',index\_col = 'Date', parse\_dates = **True**)  
CO2\_df

Out[12]:

|  | **Unnamed: 0** | **co2** | **NG\_F** | **coal-f** | **NG\_S** | **crude** | **co2\_diff** | **crude\_diff** | **crude\_diff\_3** | **NG\_F\_5** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** |  |  |  |  |  |  |  |  |  |  |
| **2000-08-28** | 0 | 367.70 | 4.835 | NaN | 4.63 | 32.869999 | -0.38 | NaN | NaN | NaN |
| **2000-09-04** | 1 | 367.32 | 4.880 | NaN | 4.66 | 32.869999 | -0.38 | 0.000000 | NaN | NaN |
| **2000-09-11** | 2 | 367.33 | 5.205 | NaN | 4.82 | 35.099998 | 0.01 | 2.229999 | NaN | NaN |
| **2000-09-18** | 3 | 366.86 | 5.131 | NaN | 5.05 | 36.880001 | -0.47 | 1.780003 | NaN | NaN |
| **2000-09-25** | 4 | 366.86 | 5.200 | NaN | 5.17 | 31.570000 | 0.00 | -5.310001 | 0.000000 | NaN |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2021-01-18** | 1064 | 415.42 | 2.446 | 67.550003 | 2.54 | 52.905001 | 0.35 | 0.045001 | 0.508333 | 2.700 |
| **2021-01-25** | 1065 | 416.01 | 2.564 | 67.800003 | 2.71 | 52.690001 | 0.59 | -0.215000 | 2.115001 | 2.518 |
| **2021-01-25** | 1066 | 416.01 | 2.564 | 66.500000 | 3.12 | 61.620441 | 0.00 | 8.930441 | 2.609999 | 2.539 |
| **2021-01-25** | 1067 | 416.01 | 2.564 | 67.000000 | 3.12 | 61.620441 | 0.00 | 0.000000 | 0.045001 | 2.700 |
| **2021-01-25** | 1068 | 416.01 | 2.564 | 66.000000 | 3.12 | 61.620441 | 0.00 | 0.000000 | -0.215000 | 2.737 |

1069 rows × 10 columns

In [14]:

pacf = plot\_pacf(CO2\_df['co2'], lags = 10)  
plt.title('PACF')  
pacf.show()

<ipython-input-14-85c01b02a8f2>:3: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend\_inline, which is a non-GUI backend, so cannot show the figure.  
 pacf.show()

In [15]:

acf = plot\_acf(CO2\_df['co2'], lags = 10)  
plt.title('ACF')  
pacf.show()

<ipython-input-15-3d6aab93ef7d>:3: UserWarning: Matplotlib is currently using module://ipykernel.pylab.backend\_inline, which is a non-GUI backend, so cannot show the figure.  
 pacf.show()

Explanation:

We use PACF plot and ACF plot with 10 lags to determine the order number of auto-corrrelation of CO2 concentration.

According to the PACF plot, the first-order is high, and the coefficients after that converge to 0 very fast.

According to the ACF plot, within 10 weeks, there is no obvious evidence showing that the coefficients converge to 0.

Hence, we should use the PACF plot to determine the orrder number of auto-correlation of gasoline price and it is AR(1).

In [ ]: