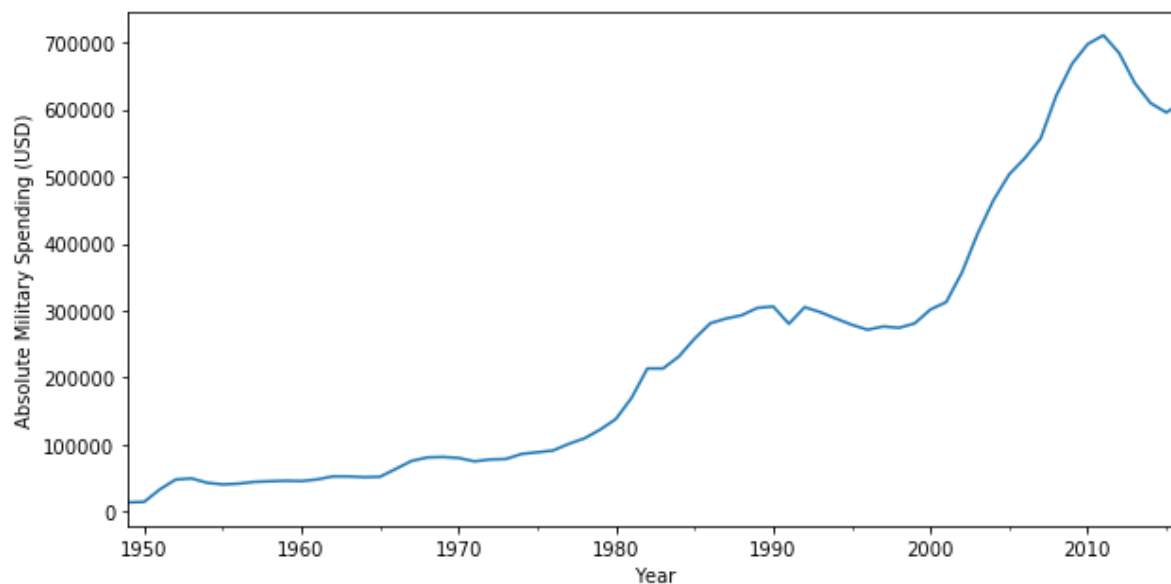


```
In [34]: from pandas import Series
import numpy as np
from matplotlib import pyplot as plt
from statsmodels.tsa.stattools import acf, pacf
import statsmodels.tsa.stattools as ts
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima_model import ARIMA
from scipy import stats
import scipy
import pandas as pd
import math
from statsmodels.graphics.api import qqplot

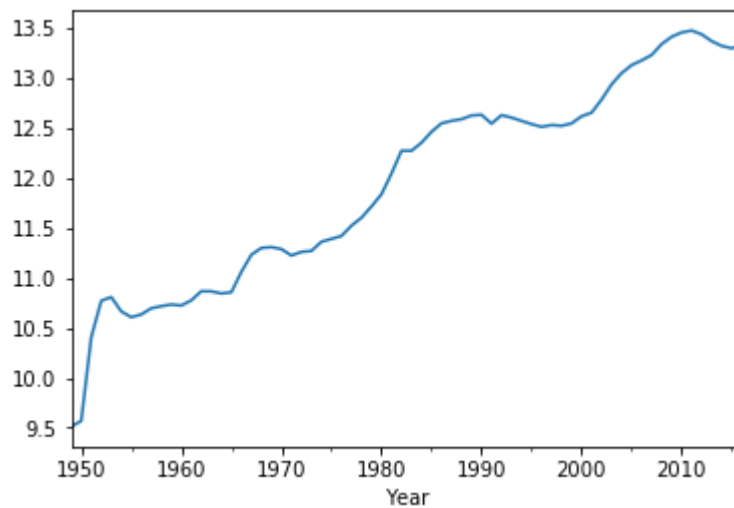
from sklearn.metrics import mean_squared_error
from numpy import std, mean, sqrt
from statsmodels.stats import diagnostic as diag
```

## Exploratory Analysis

```
In [35]: #plotting the raw time series data
series = Series.from_csv('military.csv', header=0)
plt.ylabel('Absolute Military Spending (USD)')
series.plot(figsize=(10,5))
plt.show()
```



```
In [36]: #Log-form manipulation to inform on relative changes and improve the efficiency of the model.  
series = series.apply(np.log)  
series.plot()  
plt.show()
```

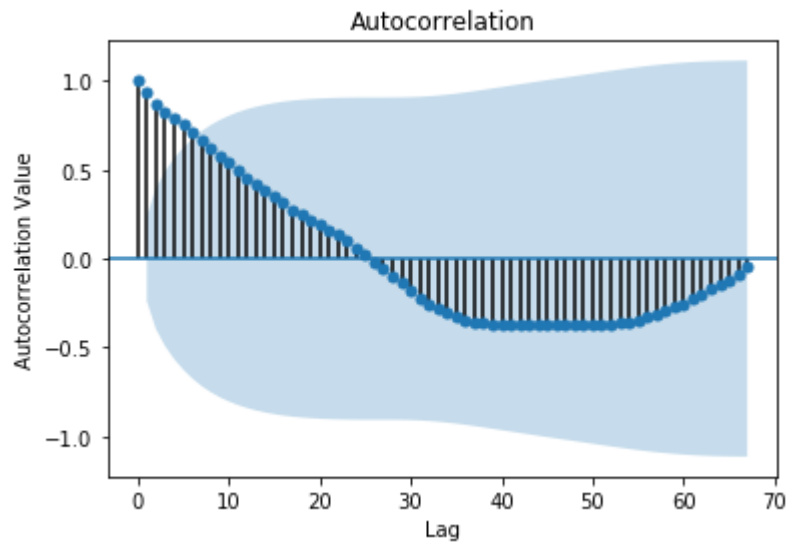


## Stationarity Tests

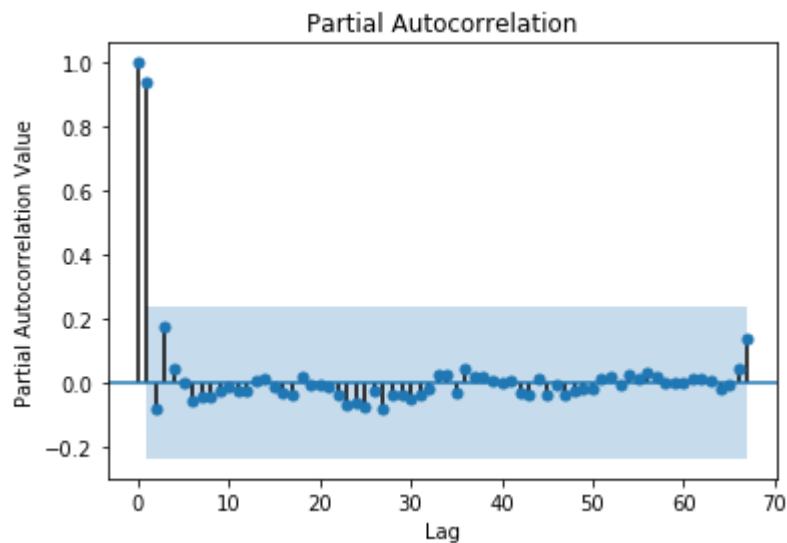
```
In [37]: #ACF and PACF tests for stationarity.
#The highlighted regions depict a 95% confidence interval,
#under the assumption that the time series is iid, suggesting that correlation
#values outside of this cone are very likely a correlation and not a statistic
#al fluke.
#Confidence intervals for ACF are drawn as a cone.

plot_acf(series)
plt.xlabel('Lag')
plt.ylabel('Autocorrelation Value')
plt.show()

#steady decrease downwards
## If one or more lags pierce those dashed lines, then the lag(s) is significant
ntly different from zero and the series is not white noise.
```



```
In [38]: plot_pacf(series)
plt.xlabel('Lag')
plt.ylabel('Partial Autocorrelation Value')
plt.show()
```



## Auamented Dickey-Fuller Test

```
In [21]: result = ts.adfuller(x, 2, regression='c') #maximum lag which is included in test
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
ADF Statistic: -0.546798
p-value: 0.882556
Critical Values:
    1%: -3.535
    5%: -2.907
   10%: -2.591
```

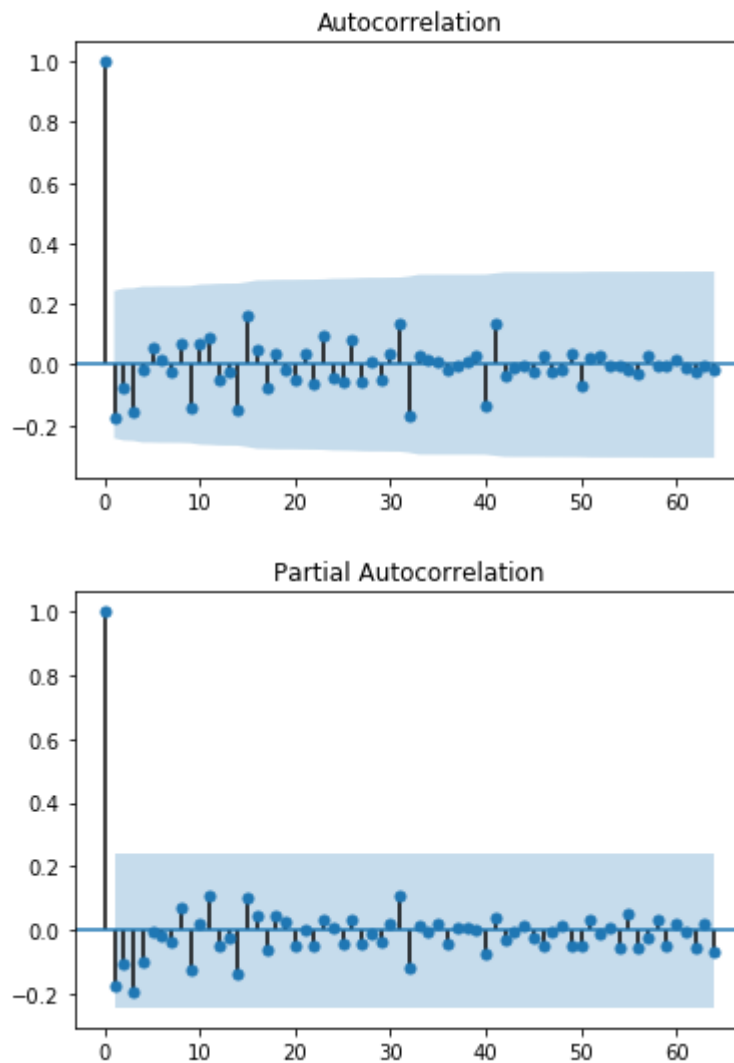
```
In [98]: x_diff = np.diff(x)
x_diff = np.diff(x_diff)
x_diff = np.diff(x_diff)
result = ts.adfuller(x_diff, 2, regression='c') #maximum lag which is included in test
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
ADF Statistic: -9.102921
p-value: 0.000000
Critical Values:
    1%: -3.541
    5%: -2.909
   10%: -2.592
```

```
In [108]: # acf and pacf
plot_acf(x_diff)
plot_pacf(x_diff)

# hist plot
count, division = np.histogram(x_diff)
plt.show()

#histogram
count, division = np.histogram(x_diff)
plt.show()
```



## Practical signifiance

```
In [77]: #Practical significance for difference in means between training and test data.

size = len(x_diff)//2
train, test= x_diff[:size], x_diff[size:]

def cohen_d(x,y):
    nx = len(x)
    ny = len(y)
    dof = nx + ny - 2
    return (mean(x) - mean(y)) / sqrt(((nx-1)*std(x, ddof=1) ** 2 + (ny-1)*std
(y, ddof=1) ** 2) / dof)

print(cohen_d(test,train))
```

0.0177790794304

## Model Identification

In [78]: *#Source: <https://stackoverflow.com/questions/30901460/non-invertible-of-a-arima-model>*

```
p_values = [0, 1, 2, 4]
d_values = range(0, 3)
q_values = range(0, 3)

pdq = []
stderr = []

for p in p_values:
    for d in d_values:
        for q in q_values:
            order = (p,d,q)
            pdq.append(order)
            history = [x for x in train]
            predictions = []
            try:
                for t in range(len(test)):
                    model = ARIMA(history, order)
                    model_fit = model.fit(dis=0)
                    output = model_fit.forecast()
                    yhat = output[0]
                    predictions.append(yhat)
                    obs = test[t]
                    history.append(obs)
                error = mean_squared_error(test, predictions)
                stderr.append(error)

            except: # ignore the error and go on

            pass
```

[illegible]



[illegible]

[illegible]

[illegible]

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:584: R
untimeWarning: overflow encountered in exp
    newparams = ((1-np.exp(-params))/
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:585: R
untimeWarning: overflow encountered in exp
    (1+np.exp(-params))).copy()
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:585: R
untimeWarning: invalid value encountered in true_divide
    (1+np.exp(-params))).copy()
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:586: R
untimeWarning: overflow encountered in exp
    tmp = ((1-np.exp(-params))/
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:587: R
untimeWarning: overflow encountered in exp
    (1+np.exp(-params))).copy()
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:587: R
untimeWarning: invalid value encountered in true_divide
    (1+np.exp(-params))).copy()
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
```

[illegible]

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:612: R
untimeWarning: divide by zero encountered in true_divide
    invarcoefs = -np.log((1-params)/(1+params))
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:654: R
untimeWarning: invalid value encountered in log
    invmacoefs = -np.log((1-macoefs)/(1+macoefs))
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: Con
vergenceWarning: Maximum Likelihood optimization failed to converge. Check ml
e_retvals
```

[illegible]

[illegible]



[illegible]

[illegible]

```
In [81]: keys = pdq
         values = stderr
         d = dict(zip(keys, values))
         print (d)
```

```
{(0, 0, 0): 2.9270637251015354e-06, (0, 0, 1): 3.9560322657175338e-06, (0, 0, 2): 2.8977320818170993e-06, (0, 1, 0): 1.1617086909392845e-05, (0, 1, 1): 3.0984474614688966e-06, (0, 1, 2): 7.7248472306088377e-06, (0, 2, 0): 3.8494757828279521e-05, (0, 2, 1): 1.2555802040464861e-05, (0, 2, 2): 3.3837622457271891e-06, (1, 0, 0): 2.5902093332326673e-06, (1, 0, 1): 6.8999283052727381e-06, (1, 0, 2): 2.0930473825749393e-05, (1, 1, 0): 3.351523538509614e-06, (1, 1, 1): 4.4089501631587672e-06, (1, 1, 2): 1.863421723412184e-05, (1, 2, 0): 1.4454719318780986e-05}
```

```
In [102]: min(d, key=d.get)
```

```
Out[102]: (1, 0, 0)
```

```
In [103]: #note that ARIMA model will fit to the differenced data
model = ARIMA(train, order=(1,0,0))
model_fit = model.fit(disp=False)
```

```
In [104]: forecast, stderr, conf = model_fit.forecast() #The forecast() function allows
the confidence interval to be specified.
#The alpha argument on the forecast() function specifies the confidence level.
It is set by default to alpha=0.05, which is a 95% confidence interval. This
is a sensible and widely used confidence interval.
#An alpha of 0.05 means that the ARIMA model will estimate the upper and lower
values around the forecast where there is a only 5% of the time the real valu
e will not be in that range.
print('Expected: %.3f' % test[0])
print('Forecast: %.3f' % forecast)
print('Standard Error: %.3f' % stderr)
intervals = [0.2, 0.1, 0.05, 0.01]
for a in intervals:
    forecast, stderr, conf = model_fit.forecast(alpha=a)
    print('%.1f%% Confidence Interval: %.3f between %.3f and %.3f' % ((1-a)
)*100, forecast, conf[0][0], conf[0][1]))
```

Expected: 0.003

Forecast: 0.003

Standard Error: 0.008

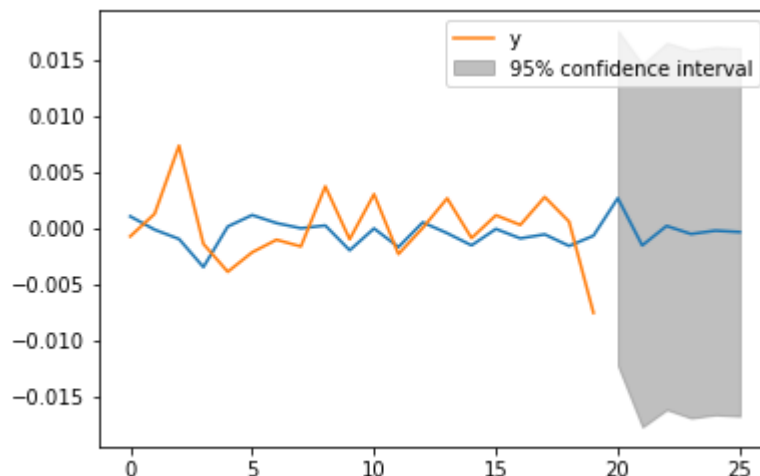
80.0% Confidence Interval: 0.003 between -0.007 and 0.012

90.0% Confidence Interval: 0.003 between -0.010 and 0.015

95.0% Confidence Interval: 0.003 between -0.012 and 0.018

99.0% Confidence Interval: 0.003 between -0.017 and 0.022

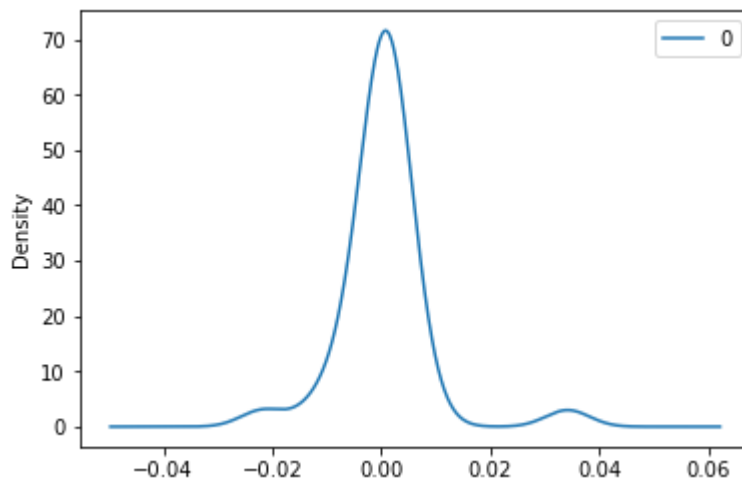
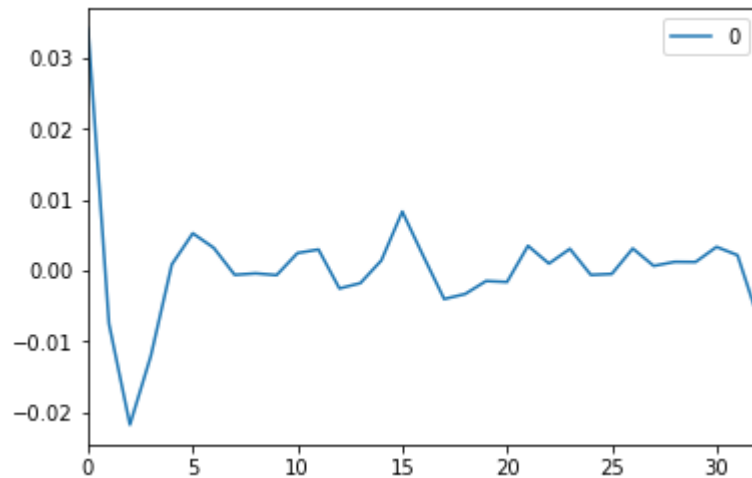
```
In [107]: model_fit.plot_predict(len(train)-20, len(train)+5)
plt.show()
```



## Model Diagnostics

```
In [109]: # plot residual erros

residuals = pd.DataFrame(model_fit.resid)
residuals.plot()
residuals.plot(kind='kde')
plt.show()
print(residuals.describe())
```



```
0
count    33.000000
mean      0.000427
std       0.008103
min      -0.021742
25%      -0.001621
50%       0.000853
75%       0.002958
max       0.034111
```