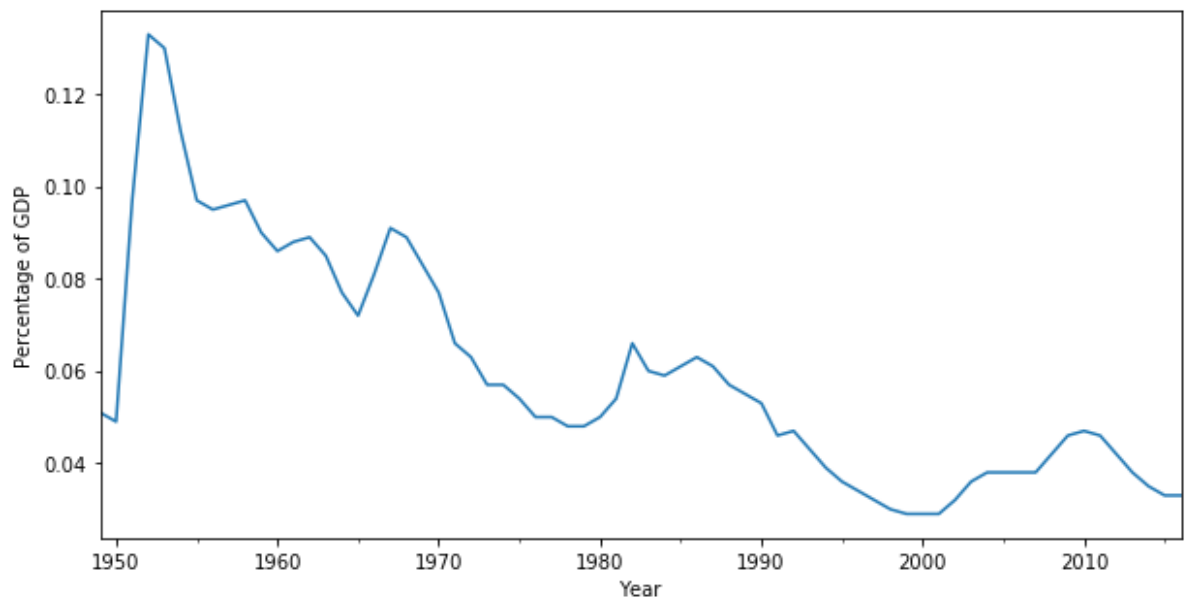


```
In [50]: from pandas import Series
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from statsmodels.tsa.stattools import acf, pacf
import statsmodels.tsa.stattools as ts
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_squared_error
from statsmodels.stats import diagnostic as diag
from scipy import stats
import scipy
import math
from statsmodels.graphics.api import qqplot
from numpy import std, mean, sqrt
```

## Exploratory Analysis

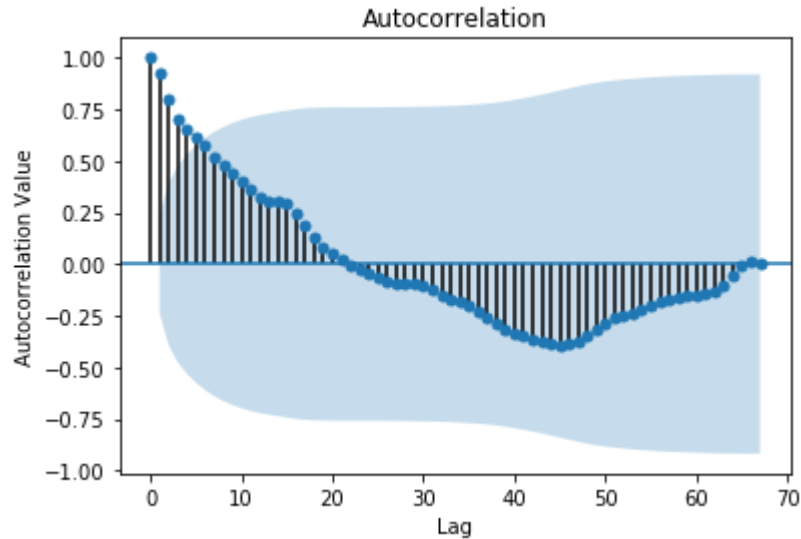
```
In [51]: series = Series.from_csv('gdp.csv', header=0)
plt.ylabel('Percentage of GDP')
series.plot(figsize=(10,5))
plt.show()
```



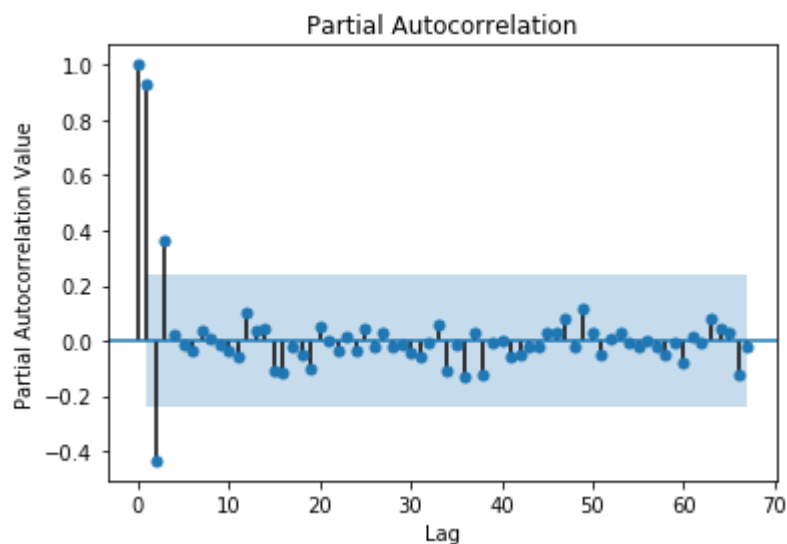
## Stationarity Tests

```
In [52]: #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 statistical
#fluke.
#Confidence intervals for ACF are drawn as a cone.

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



```
In [53]: plot_pacf(series)
plt.xlabel('Lag')
plt.ylabel('Partial Autocorrelation Value')
plt.show()
```



## Augmented Dickey-Fuller Test

```
In [56]: result = ts.adfuller(x,1, regression = 'c')
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: -2.302020
p-value: 0.171317
Critical Values:
    1%: -3.534
    5%: -2.906
   10%: -2.591
```

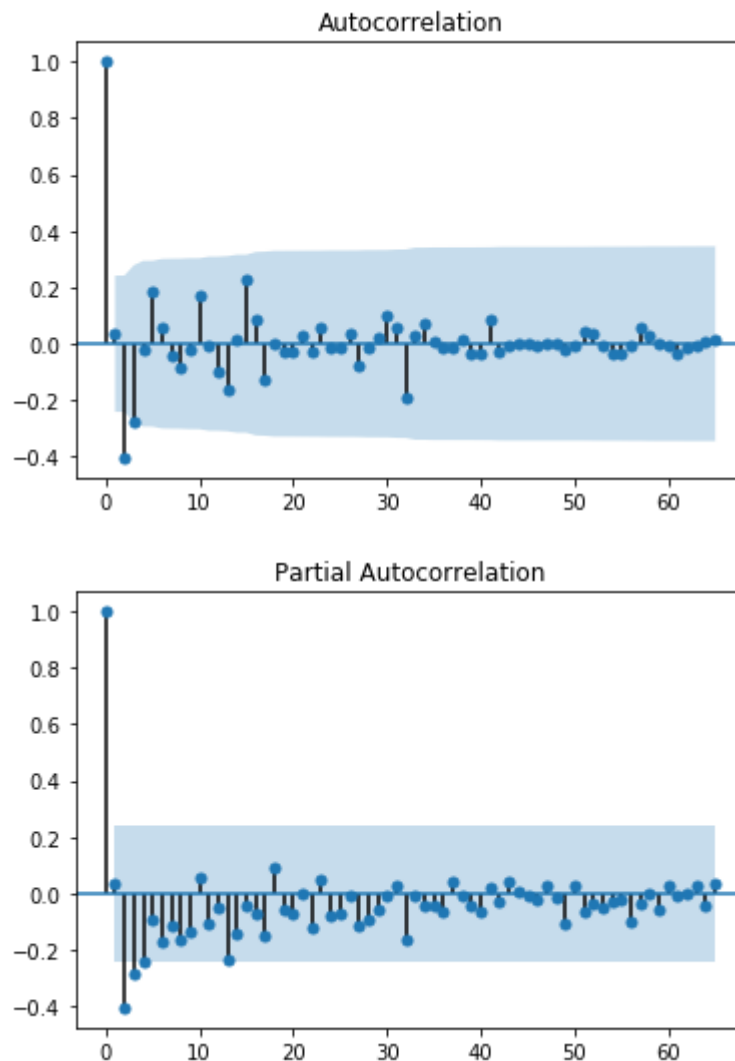
```
In [60]: x_diff = np.diff(x)
x_diff = np.diff(x_diff)
result = ts.adfuller(x_diff,1,regression='c') #maximum lag which is included i
n test
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
ADF Statistic: -9.425920
p-value: 0.000000
Critical Values:
    1%: -3.537
    5%: -2.908
   10%: -2.591
```

```
In [99]: # 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 Significance

```
In [67]: size = len(x_diff)//2
train, test= x_diff[:size], x_diff[size:]

#test for practical significance

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))

#

0.0315758778594
```

## Model Identification

```

In [68]: #finding optimal parameters for ARIMA model

"""Split the dataset into training and test sets.
Walk the time steps in the test dataset.
Train an ARIMA model.
Make a one-step prediction.
Store prediction; get and store actual observation.
Calculate error score for predictions compared to expected values."""

#Source: https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/

# evaluate an ARIMA model for a given order (p,d,q)
def evaluate_arima_model(X, arima_order):
    size = len(X)//2
    # prepare training dataset
    train, test= X[0:size], X[size:]
    history = [x for x in train]
    # make predictions
    predictions = list()
    for t in range(len(test)):
        model = ARIMA(history, order=arima_order)
        model_fit = model.fit(dis=0)
        yhat = model_fit.forecast()[0]
        predictions.append(yhat)
        history.append(test[t])
        # calculate out of sample error
    error = mean_squared_error(test, predictions)
    return error

def evaluate_models(dataset, p_values, d_values, q_values):
    dataset = dataset.astype('float32')
    best_score, best_cfg = float("inf"), None
    for p in p_values:
        for d in d_values:
            for q in q_values:
                order = (p,d,q)
                try:
                    mse = evaluate_arima_model(dataset, order)
                    if mse < best_score:
                        best_score, best_cfg = mse, order
                        print('ARIMA%s MSE=%.3f' % (order,mse))
                except:
                    continue
    print('Best ARIMA%s MSE=%.3f' % (best_cfg, best_score))

p_values = [0, 1, 2, 4, 6, 8, 10]
d_values = range(0,3)
q_values = range(0, 3)

evaluate_models(series.values, p_values, d_values, q_values)

```

```
ARIMA(0, 0, 0) MSE=0.001
ARIMA(0, 0, 1) MSE=0.000
ARIMA(0, 1, 0) MSE=0.000
ARIMA(0, 1, 1) MSE=0.000
ARIMA(0, 2, 0) MSE=0.000
ARIMA(0, 2, 1) MSE=0.000
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:473: HessianInversionWarning: Inverting hessian failed, no bse or cov_params available
```

```
    'available', HessianInversionWarning)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:628: RuntimeWarning: overflow encountered in exp
```

```
    newparams = ((1-np.exp(-params))/(1+np.exp(-params))).copy()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:628: RuntimeWarning: invalid value encountered in true_divide
```

```
    newparams = ((1-np.exp(-params))/(1+np.exp(-params))).copy()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:629: RuntimeWarning: overflow encountered in exp
```

```
    tmp = ((1-np.exp(-params))/(1+np.exp(-params))).copy()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:629: RuntimeWarning: invalid value encountered in true_divide
```

```
    tmp = ((1-np.exp(-params))/(1+np.exp(-params))).copy()
```

```
ARIMA(0, 2, 2) MSE=0.000
ARIMA(1, 0, 0) MSE=0.000
ARIMA(1, 0, 1) MSE=0.000
ARIMA(1, 0, 2) MSE=0.000
ARIMA(1, 1, 0) MSE=0.000
ARIMA(1, 2, 0) MSE=0.000
ARIMA(2, 0, 0) MSE=0.000
ARIMA(2, 0, 1) MSE=0.000
```

```

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)
ARIMA(2, 1, 1) MSE=0.000

```



[illegible]

```
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  "Check mle_retvals", ConvergenceWarning)
```

```
ARIMA(2, 2, 0) MSE=0.000
```

```
ARIMA(2, 2, 1) MSE=1.675
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

ARIMA(6, 0, 1) MSE=0.000

[illegible]

```
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:654:
RuntimeWarning: divide by zero encountered in true_divide
    invmacoefs = -np.log((1-macoefs)/(1+macoefs))
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
```

```
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:652:
RuntimeWarning: invalid value encountered in double_scalars
    tmp[kiter] = (macoefts[kiter]-b *macoefts[j-kiter-1])/(1-b**2)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
```

```
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\regression\linear_mo
del.py:1127: RuntimeWarning: invalid value encountered in sqrt
    return rho, np.sqrt(sigmasq)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
    "Check mle_retvals", ConvergenceWarning)
```

```

"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\base\model.py:496: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
k mle_retvals
"Check mle_retvals", ConvergenceWarning)

```

Best ARIMA(0, 1, 0) MSE=0.000

```

C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tsa\tsatools.py:612: R
untimeWarning: invalid value encountered in log
invarcoefs = -np.log((1-params)/(1+params))

```

```

In [90]: model = ARIMA(test, order=(0,1,0))
model_fit = model.fit(dispatch=False)

```

```

In [93]: 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: 0.005

Forecast: 0.002

Standard Error: 0.004

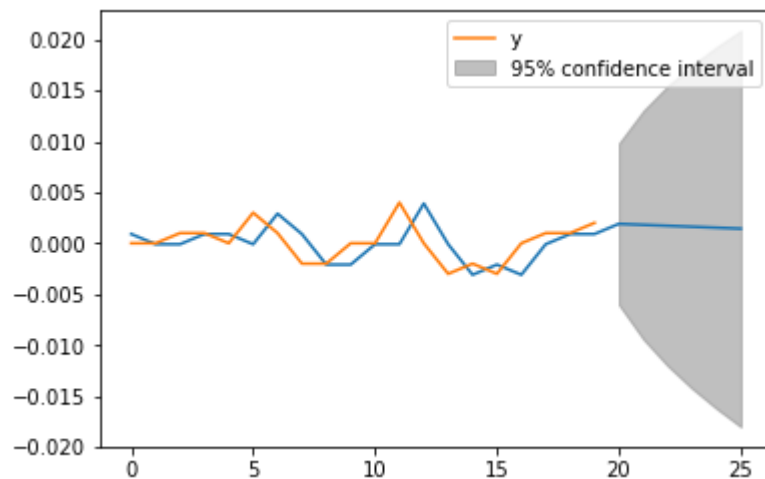
80.0% Confidence Interval: 0.002 between -0.003 and 0.007

90.0% Confidence Interval: 0.002 between -0.005 and 0.009

95.0% Confidence Interval: 0.002 between -0.006 and 0.010

99.0% Confidence Interval: 0.002 between -0.009 and 0.012

```
In [98]: model_fit.plot_predict(len(train)-20, len(train)+5)  
plt.show()
```



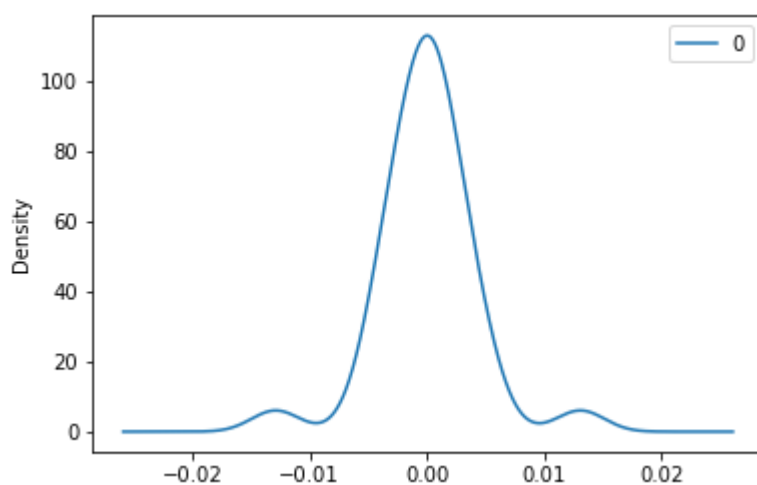
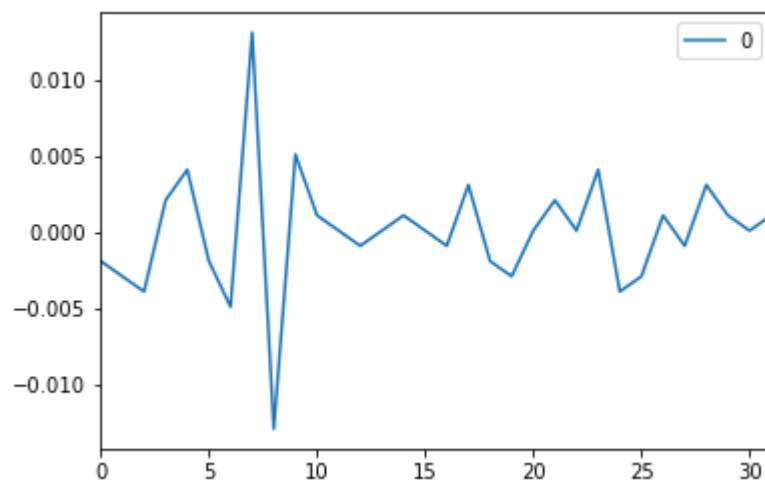
## Model Diagnostics

```
In [95]: #After fitting the model, we should check whether the model is appropriate.  
#Let's analyze the residuals and investigate the autocorrelation of the ARIMA  
model  
residuals = pd.DataFrame(model_fit.resid)
```



```
In [100]: # plot residual erros

residuals.plot()
residuals.plot(kind='kde')
plt.show()
print(residuals.describe())
```



```
0
count    3.200000e+01
mean      8.321877e-11
std       4.114171e-03
min      -1.290624e-02
25%      -1.906255e-03
50%       9.374924e-05
75%       1.343755e-03
max       1.309375e-02
```