Chapter 19: Demo Time Series với Holtwinters

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.holtwinters import ExponentialSmoothing
```

Đọc dữ liệu, kiểm tra/định dạng thời gian

In [2]:
```
df = pd.read_csv('international-airline-passengers.csv',
                 parse_dates=['Month'],
                 index_col='Month')
```

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 144 entries, 1949-01-01 to 1960-12-01
Data columns (total 1 columns):
passengers_in_thousands    144 non-null int64
dtypes: int64(1)
memory usage: 2.2 KB
```

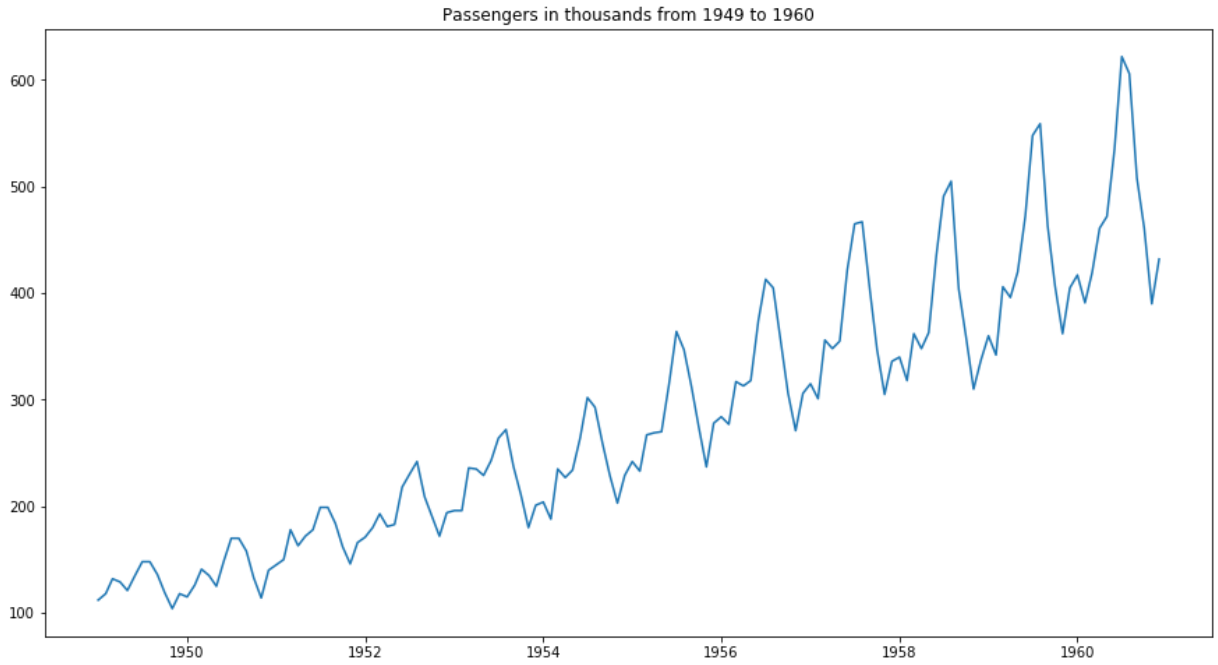
In [4]:
```
# https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html
# freq='H', 'D', 'W', 'M', 'MS': Hour, Day, Week, Month, Calendar month begin
df.index.freq = 'MS'
```

In [5]:
```
df.head()
```

Out[5]:

Month	passengers_in_thousands
1949-01-01	112
1949-02-01	118
1949-03-01	132
1949-04-01	129
1949-05-01	121

In [6]:
```
plt.figure(figsize=(15,8))
plt.plot(df)
plt.title("Passengers in thousands from 1949 to 1960")
plt.show()
```

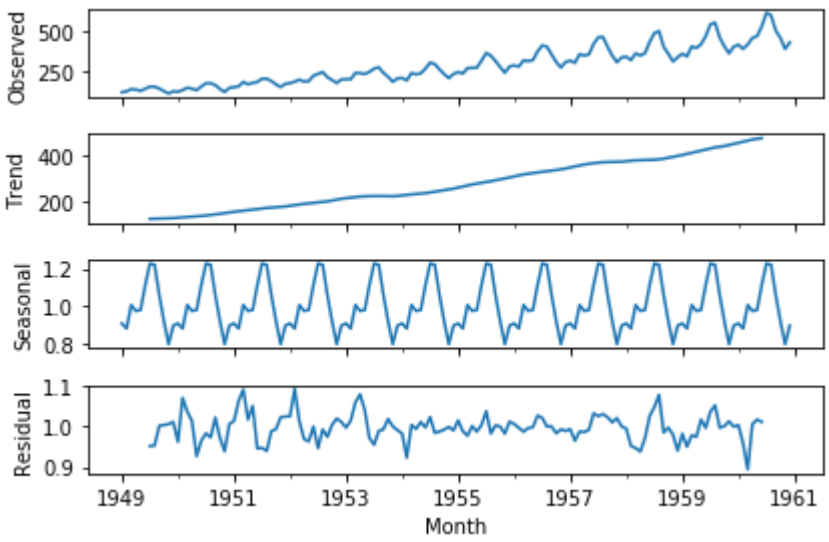


Decomposition

In [7]:
```
from statsmodels.tsa.seasonal import seasonal_decompose
result = seasonal_decompose(df, model='multiplicative')
result
```

Out[7]:
```
<statsmodels.tsa.seasonal.DecomposeResult at 0x1ebeedb6710>
```

In [8]:
```
result.plot()
plt.show()
```



Chia dữ liệu train/test => Áp dụng mô hình

In [9]:
```
train, test = df.iloc[:130, 0], df.iloc[130:, 0]
```

In [10]:
```
train[0:5]
```

Out[10]:
```
Month
1949-01-01    112
1949-02-01    118
1949-03-01    132
1949-04-01    129
1949-05-01    121
Freq: MS, Name: passengers_in_thousands, dtype: int64
```

In [11]:
```
test[0:5]
```

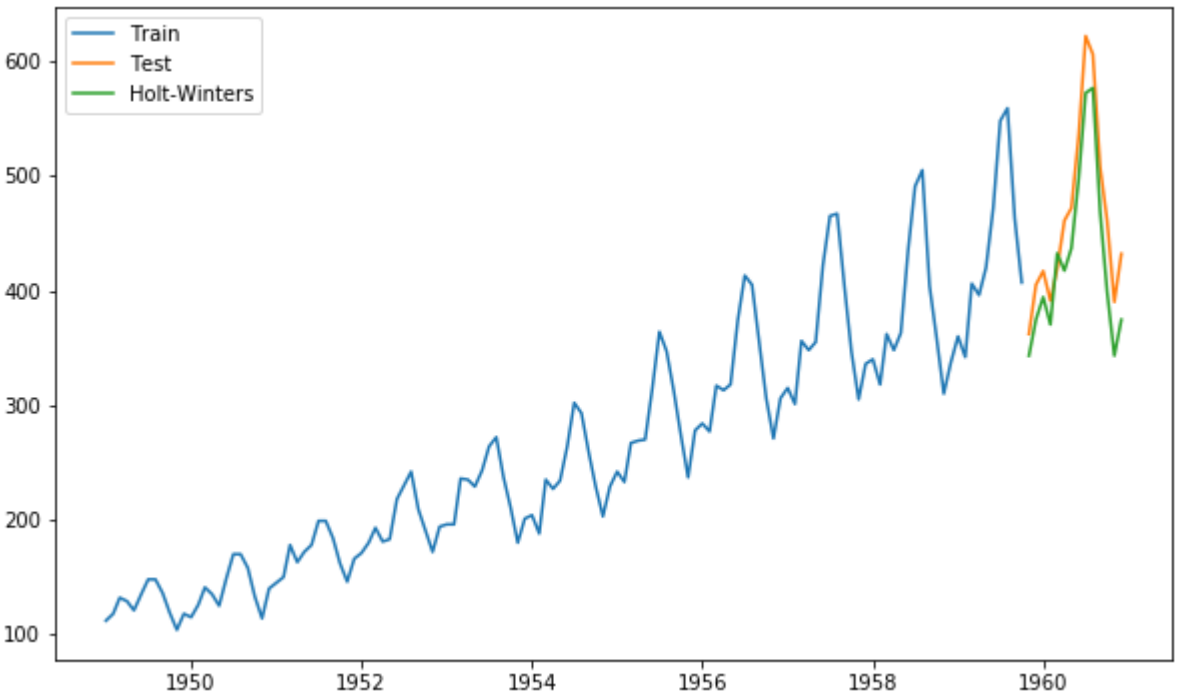
Out[11]:
```
Month
1959-11-01    362
1959-12-01    405
1960-01-01    417
1960-02-01    391
1960-03-01    419
Freq: MS, Name: passengers_in_thousands, dtype: int64
```

- https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing (https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing
- https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.p (https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.

In [12]:
```
model = ExponentialSmoothing(train, seasonal='mul',
                             seasonal_periods=12).fit()
pred = model.predict(start=test.index[0],
                     end=test.index[-1])
```

In [13]:
```
plt.figure(figsize=(10,6))
plt.plot(train.index, train, label='Train')
plt.plot(test.index, test, label='Test')
plt.plot(pred.index, pred, label='Holt-Winters')
plt.legend(loc='best')
```

Out[13]:
```
<matplotlib.legend.Legend at 0x1ebf244f2b0>
```



Dự đoán

In [14]:
```
import datetime
s = datetime.datetime(1961, 1, 1)
e = datetime.datetime(1961, 12, 1)
pred_next_12_month = model.predict(start= s, end=e)
pred_next_12_month
```

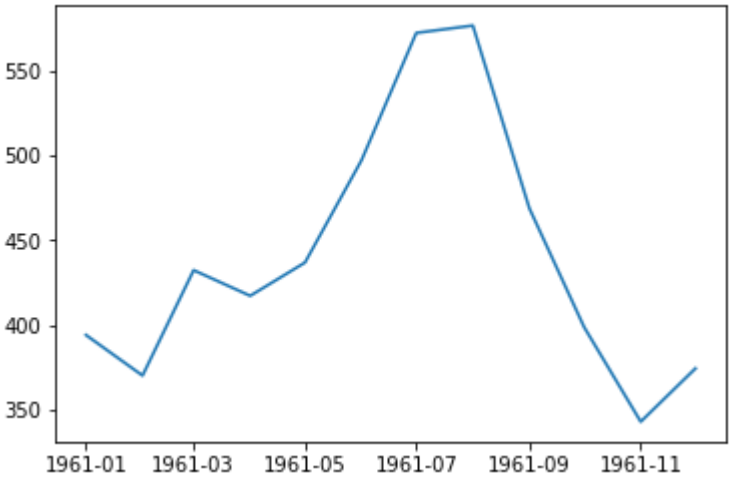
Out[14]:
```
1961-01-01    394.217184
1961-02-01    370.307412
1961-03-01    432.436334
1961-04-01    417.324009
1961-05-01    436.929699
1961-06-01    497.258023
1961-07-01    572.320206
1961-08-01    576.655572
1961-09-01    468.866230
1961-10-01    398.843314
1961-11-01    343.119846
1961-12-01    374.536786
Freq: MS, dtype: float64
```

In [15]:
```
x = pd.Series(pred_next_12_month)
type(x)
```

Out[15]:
```
pandas.core.series.Series
```

In [16]:
```
plt.plot(x.index, x.values)
```

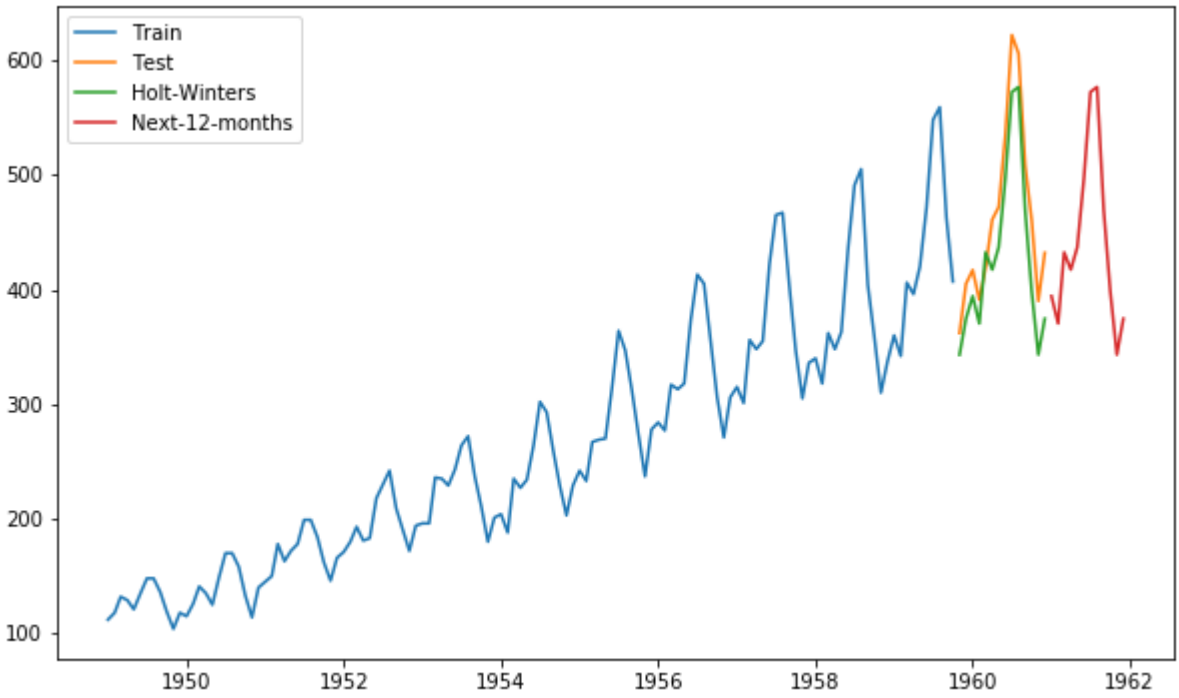
Out[16]:
```
[<matplotlib.lines.Line2D at 0x1ebf2419e80>]
```



Trực quan hóa dữ liệu

In [17]:
```
plt.figure(figsize=(10,6))
plt.plot(train.index, train, label='Train')
plt.plot(test.index, test, label='Test')
plt.plot(pred.index, pred, label='Holt-Winters')
plt.plot(x.index, x.values, label='Next-12-months')
plt.legend(loc='best')
```

Out[17]:
```
<matplotlib.legend.Legend at 0x1ebf24dc5c0>
```



In [ ]:
```

```