KPSS Test for Stationarity

by Selva Prabhakaran (https://www.machinelearningplus.com/author/selva86/) |

(/#facebook) (/#twitter) (/#whatsapp) (/#linkedin) (/#reddit)
(/#google_bookmarks)
+
(https://www.addtoany.com/share#url=https%3A%2F%2Fwww.machinelearningplus.com%2Ftime-series%2Fkpss-test-for-stationarity%2F&title=KPSS%20Test%20for%20Stationarity)

KPSS test is a statistical test to check for stationarity of a series around a deterministic trend. Like ADF test, the KPSS test is also commonly used to analyse the stationarity of a series. However, it has couple of key differences compared to the ADF test in function and in practical usage. Therefore, is not safe to just use them interchangeably. We'll discuss this detail with simplified examples.

Recommended

Time Series Analysis in Python – A Comprehensive Guide with Examples (https://www.machinelearningplus.com/time-series/time-series-analysis-python/)

Vector Autoregression (VAR) – Comprehensive Guide with Examples in Python (https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/)

ARIMA Model – Complete Guide to Time Series Forecasting in Python (https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/)

Augmented Dickey Fuller Test (ADF Test) – Must Read Guide (https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/)

KPSS Test for Stationarity (https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/)

1. Introduction

The KPSS test, short for, Kwiatkowski-Phillips-Schmidt-Shin (KPSS), is a type of Unit root test that tests for the stationarity of a given series around a deterministic trend.

In other words, the test is somewhat similar in spirit with the ADF test.

A common misconception, however, is that it can be used interchangeably with the ADF test (). This can lead to misinterpretations about the stationarity, which can easily go undetected causing more problems down the line.

In this article, you will see how to implement KPSS test in python, how it is different from ADF test and when and what all things you need to take care when implementing a KPSS test.

ezoic (https://www.ezoic.com/what-is-ezoic/)
report this ad

Top Posts & Pages

ARIMA Model - Complete Guide to Time Series Forecasting in Python (https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/)

Parallel Processing in Python - A Practical Guide with Examples (https://www.machinelearningplus.com/python/parallel-processing-python/)

Let's begin.

2. How to implement KPSS test

In python, the `statsmodel` package provides a convenient implementation of the KPSS test.

A key difference from ADF test is the null hypothesis of the KPSS test is that the series is stationary.

So practically, the interpretaion of p-value (https://www.machinelearningplus.com/statistics/p-value/) is just the opposite to each other.

Turing
U.S Software Developers Jobs. Work Remotely.
No Visa needed. Full time jobs. High salary.
Sign Up Now

That is, if p-value is < signif level (say 0.05), then the series is non-stationary. Whereas in ADF test, it would mean the tested series is stationary.

Alright let's implement the test on the 'daily temperatures' dataset.

This dataset contains the minimum daily temperature of the Melbourne city from 1981 to 1990. Let's first read it in pandas and plot it.

```
# Import data
import pandas as pd
import numpy as np
path = 'https://raw.githubusercontent.com/selva86/datasets/master/daily-min-temperatures.csv (https://
df = pd.read_csv(path, parse_dates=['Date'], index_col='Date')
df.plot(title='Daily Temperatures', figsize=(14,8), legend=None);
```

(https://www.machinelearningplus.com/wp-content/uploads/2019/11/daily_min_temperature_melbourne.png)

To implement the KPSS test, we'll use the `kpss` function from the `statsmodel`. The code below implements the test and prints out the returned outputs and interpretation from the result.

Time Series Analysis in Python - A Comprehensive Guide with Examples (https://www.machinelearningplus.com/time-series/time-series-analysis-python/)

Topic Modeling with Gensim (Python) (https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/)

Top 50 matplotlib Visualizations - The Master Plots (with full python code) (https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/)

Machine Learning Better Explained! (https://www.machinelearningplus.com/)

Cosine Similarity - Understanding the math and how it works (with python codes) (https://www.machinelearningplus.com/nlp/cosine-similarity/)

Matplotlib Histogram - How to Visualize Distributions in Python (https://www.machinelearningplus.com/plots/matplotlib-histogram-python-examples/)

Python Logging - Simplest Guide with Full Code and Examples (https://www.machinelearningplus.com/python/python-logging-guide/)

101 Pandas Exercises for Data Analysis (https://www.machinelearningplus.com/python/101-pandas-exercises-python/)

Recent Posts

Matplotlib Plotting Tutorial – Complete overview of Matplotlib library (https://www.machinelearningplus.com/plots/matplotlib-plotting-tutorial/)

How to implement Linear Regression in TensorFlow (https://www.machinelearningplus.com/deep-learning/linear-regression-tensorflow/)

Brier Score – How to measure accuracy of probablistic predictions (https://www.machinelearningplus.com/statistics/brier-score/)

Modin – How to speedup pandas by changing one line of code (https://www.machinelearningplus.com/python/modin-speedup-pandas/)

Dask – How to handle large dataframes in python using parallel computing (https://www.machinelearningplus.com/python/dask-tutorial/)

Text Summarization Approaches for NLP – Practical Guide with Generative Examples (https://www.machinelearningplus.com/nlp/text-summarization-approaches-nlp-example/)

Bias Variance Tradeoff – Clearly Explained (https://www.machinelearningplus.com/machine-learning/bias-variance-tradeoff/)

Gradient Boosting – A Concise Introduction from Scratch (https://www.machinelearningplus.com/machine-learning/gradient-boosting/)

```
# KPSS test
from statsmodels.tsa.stattools import kpss
def kpss_test(series, **kw):
    statistic, p_value, n_lags, critical_values = kpss(series, **kw)
    # Format Output
    print(f'KPSS Statistic: {statistic}')
    print(f'p-value: {p_value}')
    print(f'num lags: {n_lags}')
    print('Critial Values:')
    for key, value in critical_values.items():
        print(f'   {key} : {value}')
    print(f'Result: The series is {"not " if p_value < 0.05 else ""}stationary')

kpss_test(series)
```

Result:

```
KPSS Statistic: 0.5488001951803195
p-value: 0.030675631716144246
num lags: 10
Critial Values:
   10% : 0.347
   5% : 0.463
   2.5% : 0.574
   1% : 0.739
Result: The series is not stationary
```

3. How to interpret KPSS test results

The output of the KPSS test contains 4 things:

1. The KPSS statistic
2. p-value
3. Number of lags used by the test
4. Critical values

The p-value reported by the test is the probability score based on which you can decide whether to reject the null hypothesis or not. If the p-value is less than a predefined alpha level (typically 0.05), we reject the null hypothesis.

The KPSS statistic is the actual test statistic that is computed while performing the test. For more information no the formula, the references mentioned at the end should help.

In order to reject the null hypothesis, the test statistic should be greater than the provided critical values. If it is in fact higher than the target critical value, then that should automatically reflect in a low p-value.

That is, if the p-value is less than 0.05, the kpss statistic will be greater than the 5% critical value.

Finally, the number of lags reported is the number of lags of the series that was actually used by the model equation of the kpss test. By default, the `statsmodels kpss()` uses the 'legacy' method. In `legacy` method, `int(12 * (n / 100)**(1 / 4))` number of lags is included, where n is the length of the series.

Complete Guide to Natural Language Processing (NLP) – with Practical Examples (https://www.machinelearningplus.com/nlp/natural-language-processing-guide/)

Portfolio Optimization with Python using Efficient Frontier with Practical Examples (https://www.machinelearningplus.com/machine-learning/portfolio-optimization-python-example/)

What does Python Global Interpreter Lock – (GIL) do? (https://www.machinelearningplus.com/python/python-global-interpreter-lock-gil/)

Logistic Regression in Julia – Practical Guide with Examples (https://www.machinelearningplus.com/julia/logistic-regression-in-julia-practical-guide-with-examples/)

Tags

Chatbot (https://www.machinelearningplus.com/tag/chatbot/) Classification (https://www.machinelearningplus.com/tag/cla Confidence Interval (https://www.machinelearningplus.com/tag/confidence-interval/) dask (https://www.machinelearningplus.com/tag/dask/) data.table (https://www.machinelearningplus.com/tag/data-table/) Data Manipulation (https://www.machinelearningplus.com/tag/c manipulation/) Debugging (https://www.machinelearningplus.com/tag/debugging/) Evaluation Metrics (https://www.machinelearningplus.com/tag/evaluation-metrics/) Exercises (https://www.machinelearningplus.com/tag/exer FastText (https://www.machinelearningplus.com/tag/fasttext/) Gensim (https://www.machinelearningplus.com/tag/gensim/ HuggingFace (https://www.machinelearningplus.com/tag/huggingface/) Julia (https://www.machinelearningplus.com/tag/j Julia Packages (https://www.machinelearningplus.com/tag/julia-packages/) LDA (https://www.machinelearningplus.com/tag/lda/) Lemmatization (https://www.machinelearningplus.com/tag/lemmatizatio Linear Regression (https://www.machinelearningplus.com/tag/linear-regression/) Logistic (https://www.machinelearningplus.com/tag/logistic/) Loop (https://www.machinelearningplus.com/tag/loop/) Machine Learning (https://www.machinelearningplus

4. How is KPSS test different from ADF test

A major difference between KPSS and ADF tests is the capability of the KPSS test to check for stationarity in the 'presence of a deterministic trend'.

If you go back and read the definition of the KPSS test, it tests for stationarity of the series around a 'deterministic trend'.

What that effectively means to us is, the test may not necessarily reject the null hypothesis (that the series is stationary) even if a series is steadily increasing or decreasing.

But what is a 'deterministic trend'?

The word 'deterministic' implies the slope of the trend in the series does not change permanently. That is, even if the series goes through a shock, it tends to regain its original path.

Well, let's test out the stationarity of the ' livestock ' time series from the fpp R package. Let me import the data and visualize first.

```
import pandas as pd
import numpy as np
path = 'https://raw.githubusercontent.com/selva86/datasets/master/livestock.csv (https://raw.githubuserc
df = pd.read_csv(path, parse_dates=['date'], index_col='date')
fig, axes = plt.subplots(1,1, dpi=120)
df.plot(title='Annual Livestock', figsize=(12,7), legend=None, ax=axes);
```

(https://www.machinelearningplus.com/wp-content/uploads/2019/11/Annual_Livestorck_TimeSeries.png)

There appears to be a steady increasing trend overall. So, you could expect that this series is stationary around the trend.

Lets see.

Applying the KPSS test . .

learning/) Matplotlib (https://www.machinelearningplus.com
NLP (https://www.machinelearningplus.com
NLTK (https://www.machinelearningplus.com/tag/nltk/)
Numpy (https://www.machinelearningplus.com/tag/numpy/)
P-Value (https://www.machinelearningplus.com/tag/p-value/)
plots (https://www.machinelearningplus.com/tag/plots/)
Practice Exercise (https://www.machinelearningplus.com/tag/practice-exercise/)
Python (https://www.machinelearnir
R (https://www.machinelearningplus.com/tag
Regex (https://www.machinelearningplus.com/tag/regex/)
Regression (https://www.machinelearningplus.com/tag/regression/)
Residual Analysis (https://www.machinelearningplus.com/tag/residual-analysis/)
Significance Tests (https://www.machinelearningplus.com/tag/significa tests/) Soft Cosine Similarity (https://www.machinelearningplus.com/tag/soft-cosine-similarity/) spaCy (https://www.machinelearningplus.com/tag/s
Stationarity (https://www.machinelearningplus.com/tag/stationarity/)
Statistics (https://www.machinelearningplus.com/tag/sta
Tensorflow (https://www.machinelearningplus.com/tag/tensorflow/)
TextBlob (https://www.machinelearningplus.com/tag/textblob/)
TextSummarization (https://www.machinelearningplus.com/tag/textsummarization/)
Text Summarization (https://www.machinelearningplus.com/tag/text-summarization/) Time Series (https://www.machinelearningplus.com/tag/tin series/) Topic Modeling (https://www.machinelearningplus.com/tag/topic-modeling/) T Test (https://www.machinelearningplus.com/tag/t-test/)
Visualization (https://www.machinelearningplus.com/tag/visualization

```
# KPSS test: stationarity around a trend
series = df.loc[:, 'value'].values
kpss_test(series)
```

Result:

```
KPSS Statistic: 0.5488001951803195
p-value: 0.030675631716144246
num lags: 10
Critial Values:
   10% : 0.347
   5% : 0.463
   2.5% : 0.574
   1% : 0.739
Result: The series is not stationary
```

The test says the p-value is significant (with p_value < 0.05) and hence, you can reject the null hypothesis (series is stationary) and derive that the series is NOT stationary.

But we thought the KPSS should not have rejected the null, because of the 'stationarity around a deterministic trend'. Yes?

Well, that's probably because we used the default option of the KPSS test.

By default, it tests for stationarity around a 'mean' only.

5. KPSS test around a deterministic trend

In KPSS test, to turn ON the stationarity testing around a trend, you need to explicitly pass the `regression='ct'` parameter to the `kpss`.

```
kpss_test(series, regression='ct')
```

```
KPSS Statistic: 0.11743798430485435
p-value: 0.1
num lags: 10
Critial Values:
   10% : 0.119
   5% : 0.146
   2.5% : 0.176
   1% : 0.216
Result: The series is stationary
```

Now clearly, the p-value is so high that you cannot reject the null hypothesis. So the series is stationary.

For the curious, What happens if you test the same series using the Augmented Dickey Fuller (ADF) test ()?

Let's find out

```
# ADF test on random numbers
from statsmodels.tsa.stattools import adfuller

def adf_test(series):
    result = adfuller(series, autolag='AIC')
    print(f'ADF Statistic: {result[0]}')
    print(f'p-value: {result[1]}')
    for key, value in result[4].items():
        print('Critial Values:')
        print(f'   {key}, {value}')

series = df.loc[:, 'value'].values
adf_test(series)
```

```
ADF Statistic: -0.2383837252990996
p-value: 0.9337890318823667
Critial Values:
   1%, -3.5812576580093696
Critial Values:
   5%, -2.9267849124681518
Critial Values:
   10%, -2.6015409829867675
```

Unlike KPSS test, the null hypothesis in ADF test is the series is not stationary.

Since p-value is well above the 0.05 alpha level, you cannot reject the null hypothesis. So the series is NOT stationary according to ADF test, which is actually expected.

6. Conclusion

So overall what this means to us is, if a series is stationary according to the KPSS test by setting `regression='ct'` and is not stationary according to the ADF test, it means the series is stationary around a deterministic trend and so is fairly easy to model this series and produce fairly accurate forecasts (https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/).

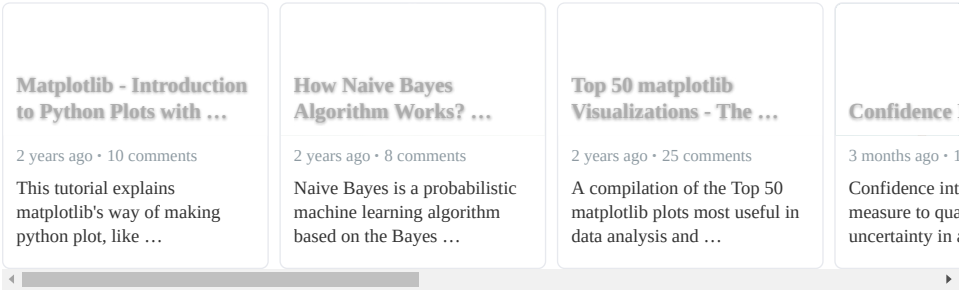
With that we have covered the core intuition behind KPSS test with practical examples. If you have any questions, please leave a line in the comments.

Thanks for joining and see you in the next one.

7. References

Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54 (1-3): 159-178.

ALSO ON **MACHINELEARNINGPLUS.COM**



What do you think?

9 Responses

Upvote Love

1 Comment **machinelearningplus.com** **Disqus' Privacy Policy** **Login**

Recommend Tweet Share **Sort by Newest**

Join the discussion…

LOG IN WITH **OR SIGN UP WITH DISQUS** ?

Name

Shruti • 2 months ago

Thank you for sharing the nice and informative article. What if ADF test result is stationary and the KPSS test with 'ct' is non-stationary for a time series?

Reply • Share ›

Subscribe **Add Disqus to your siteAdd DisqusAdd** **Do Not Sell My Data**

(https://www.ezoic.com/what-is-ezoic/)

report this ad

Search … **Search**

Subscribe to Blog

Enter your email address to receive notifications of new posts by email.

Email Address

Subscribe

Copyright ML+ (https://www.machinelearningplus.com/). **All rights reserved.**

Terms and Conditions (https://www.machinelearningplus.com/terms-of-use/)