Box Cox transform: Guerrero method

Time series decomposition

Contents

Box Cox recap

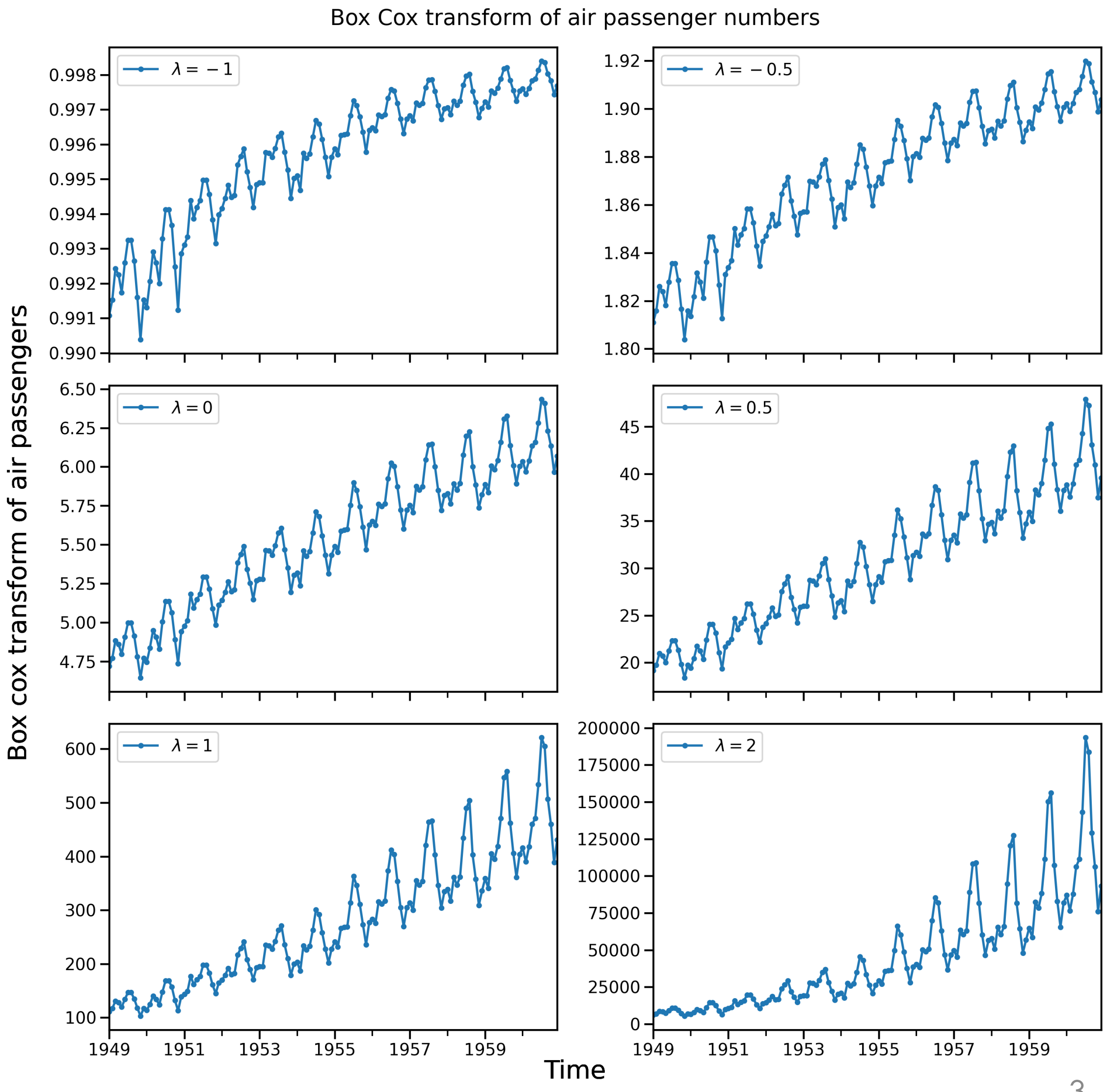
- Some forecasting & decomposition methods perform better if the variance of the time series does not change with the level of the time series (e.g., ARIMA).

- The Box Cox transform is defined as:

$$
\begin{aligned}
y^{(\lambda)} &= \frac{y^{\lambda} - 1}{\lambda}; \qquad \text{if } \lambda \neq 0 \\
&= \log(y)\,; \qquad \text{if } \lambda = 0
\end{aligned}
$$

- Different values of λ correspond to different kinds of transforms.

- How do we pick a good value for λ?



Coefficient of variation

- The coefficient of variation is a **scaled measure of variability** of a dataset.

- Coefficient of variation:

$$C_V = \frac{\text{standard deviation}}{\text{mean}} = \frac{\sigma}{\mu}$$

- It allows us to compare the variability across datasets on different scales.

Statistic	Sample 1	Sample 2
Raw data	[1, 10, 3, 50, 3, 7]	[1000, 1030, 1110, 900, 999]
Mean	12.33	1007.80
Standard deviation	17.10	67.37
Coefficient of variation	1.39	0.07

Guerrero method

Time-series analysis supported by Power Transformations

VICTOR M. GUERRERO
Instituto Tecnológico Autónomo de México, México

ABSTRACT

This paper presents some procedures aimed at helping an applied time-series analyst in the use of power transformations. Two methods are proposed for selecting a variance-stabilizing transformation and another for bias-reduction of the forecast in the original scale. Since these methods are essentially model-independent, they can be employed with practically any type of time-series model. Some comparisons are made with other methods currently available and it is shown that those proposed here are either easier to apply or are more general, with a performance similar to or better than other competing procedures.

KEY WORDS ARIMA models Bias reduction Forecasting Taylor series approximation Time-series models Variance-stabilizing

Guerrero, Victor M. "Time-series analysis supported by power transformations." *Journal of forecasting* 12.1 (1993): 37-48.

Guerrero method

- We want to pick λ so that the variance of $y_t^{(\lambda)}$ is constant.

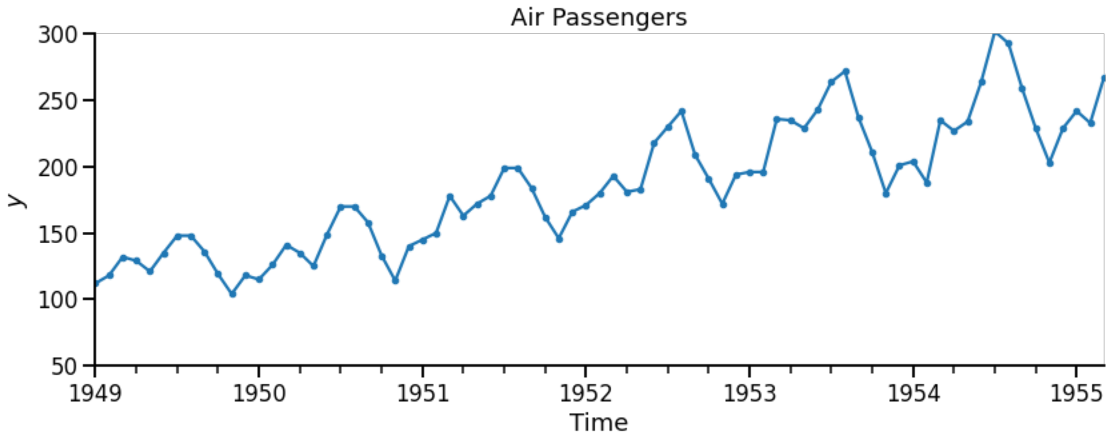
- Guerrero [1] showed that this requirement implies that:

$$\frac{\sigma[y_t]}{E[y_t]^{1-\lambda}} = constant$$

for all time steps t.

- Try a range of λ and check for which λ $\frac{\sigma[y_t]}{E[y_t]^{1-\lambda}}$ is the "most constant" over time.

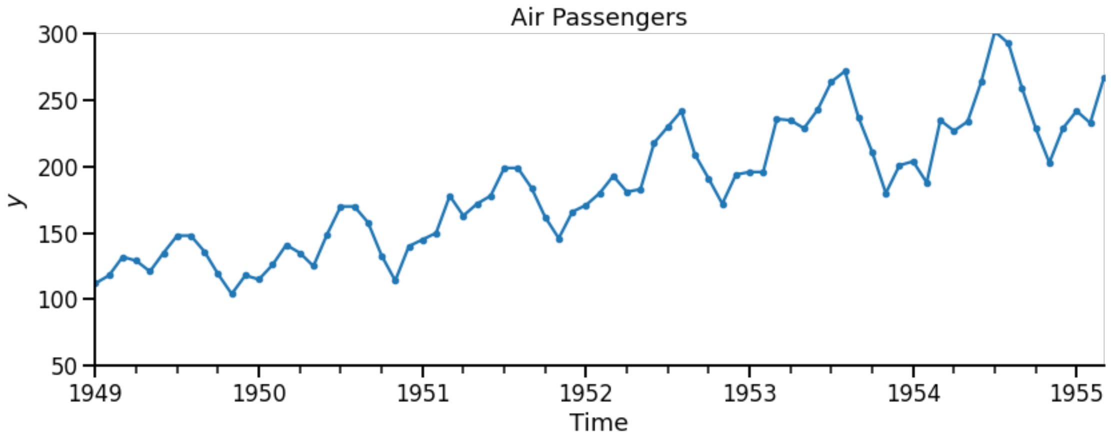
- In practice, we have one observation at each t. How do we calculate $\sigma[y_t]$ and $E[y_t]$?



[1] Guerrero, Victor M. "Time-series analysis supported by power transformations." *Journal of forecasting* 12.1 (1993): 37-48.

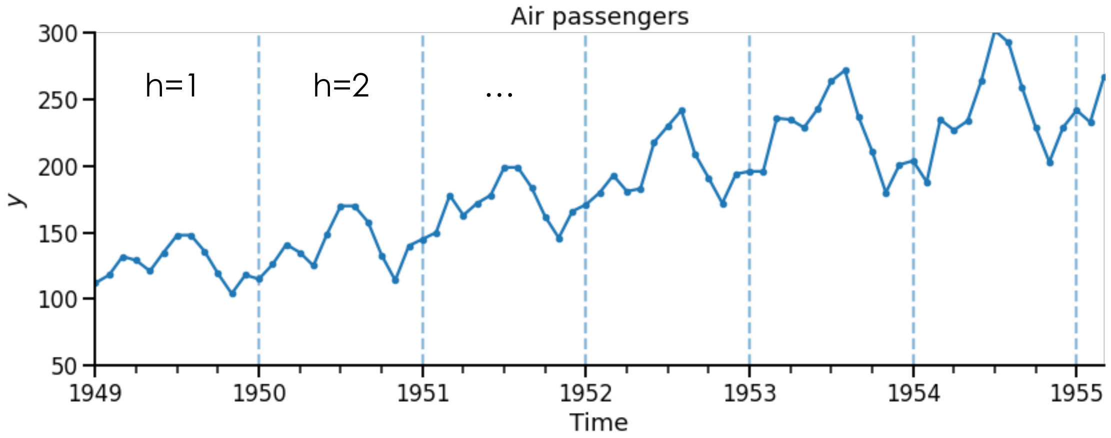
Guerrero method

- Split the time series into H evenly sized buckets (aka subseries), labelled by h.



Guerrero method

- Split the time series into H evenly sized buckets (aka subseries), labelled by h.
- Compute $\sigma[y_t]$ & $E[y_t]$ within each subseries.
- Compute $\frac{\sigma[y_t]}{E[y_t]^{1-\lambda}} = S_h(\lambda)$ for each subseries.
- How do we measure how constant $S_h(\lambda)$ is across the time series? Use the coefficient of variation of $S_h(\lambda)$ across all subseries, $C_V(\lambda)$!
- If $C_V(\lambda)$ is low, it means that $S_h(\lambda)$ is "more constant" across the time series.

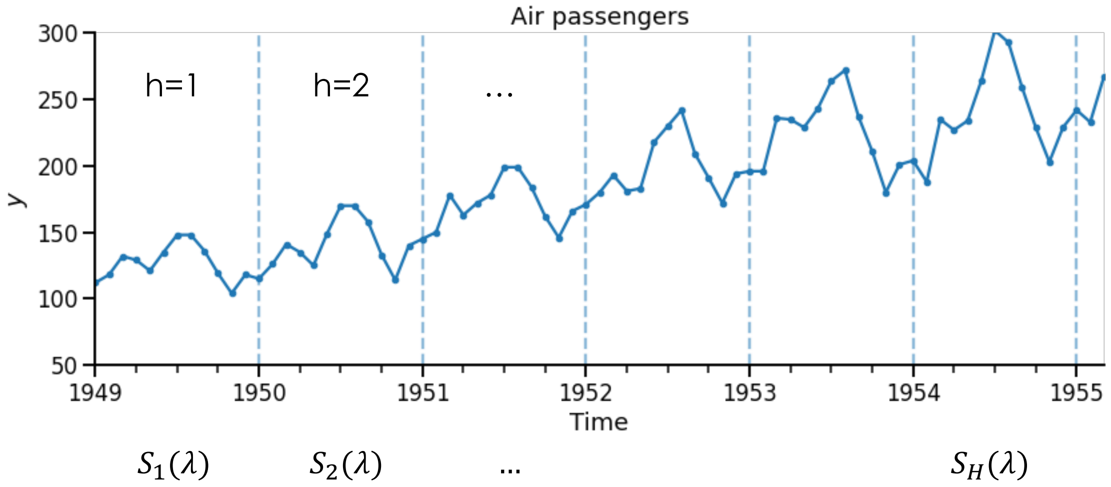


$S_1(\lambda)$ $S_2(\lambda)$ ... $S_H(\lambda)$

$$C_V(\lambda) = \frac{\sigma}{\mu} = \frac{\sigma[S_h(\lambda)]}{E[S_h(\lambda)]}$$

Guerrero method

- Compute $C_V(\lambda)$ at multiple values of λ between -5 and 5.

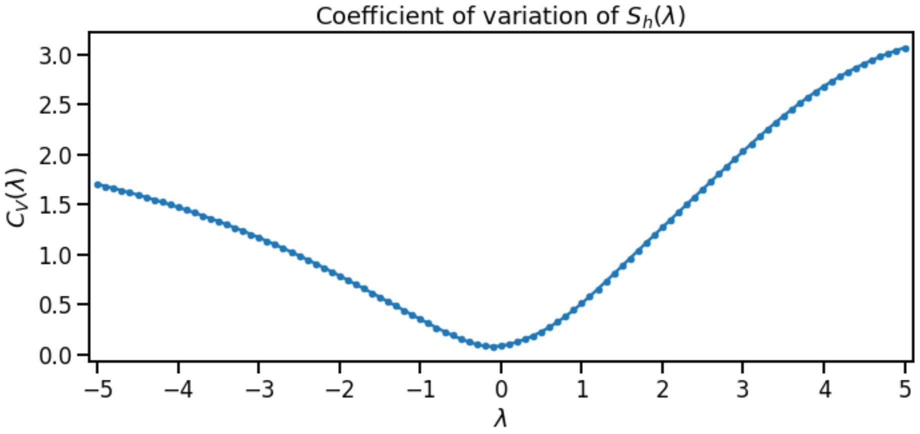


$$C_V(\lambda) = \frac{\sigma}{\mu} = \frac{\sigma[S_h(\lambda)]}{E[S_h(\lambda)]}$$

Guerrero method

- Compute $C_V(\lambda)$ at multiple values of λ between -5 and 5.

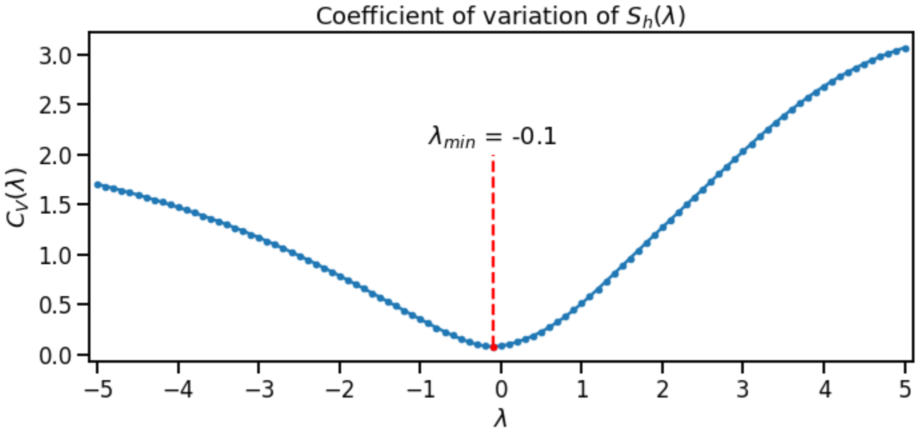
- Pick $\lambda = \lambda_{min}$ which minimizes $C_V(\lambda)$.



$$C_V(\lambda) = \frac{\sigma}{\mu} = \frac{\sigma[S_h(\lambda)]}{E[S_h(\lambda)]}$$

Guerrero method

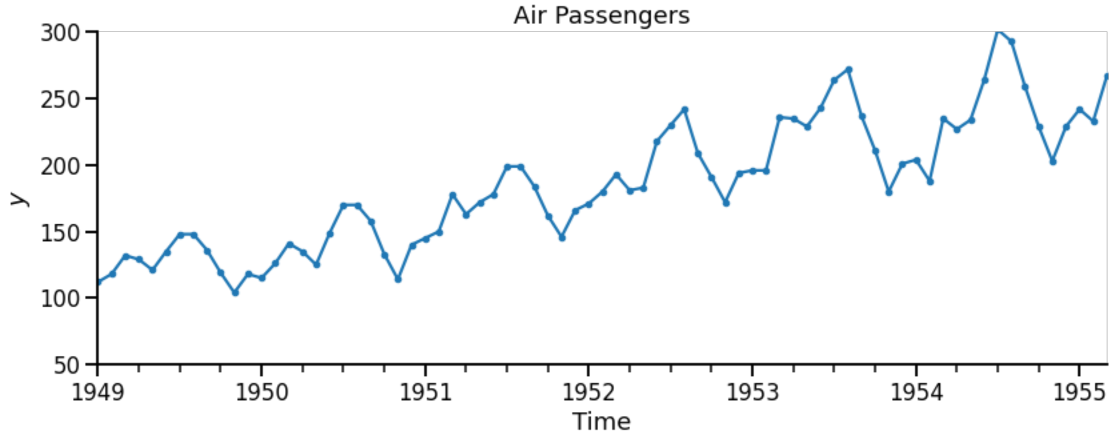
- Compute $C_V(\lambda)$ at multiple values of λ between -5 and 5.
- Pick $\lambda = \lambda_{min}$ which minimizes $C_V(\lambda)$.
- This value of λ creates a time series where $\frac{\sigma[y_t]}{E[y_t]^{1-\lambda}}$ is the "most constant" across time.
- Which implies it's the best λ to use to cause the variance of $y_t^{(\lambda)}$ to be constant.

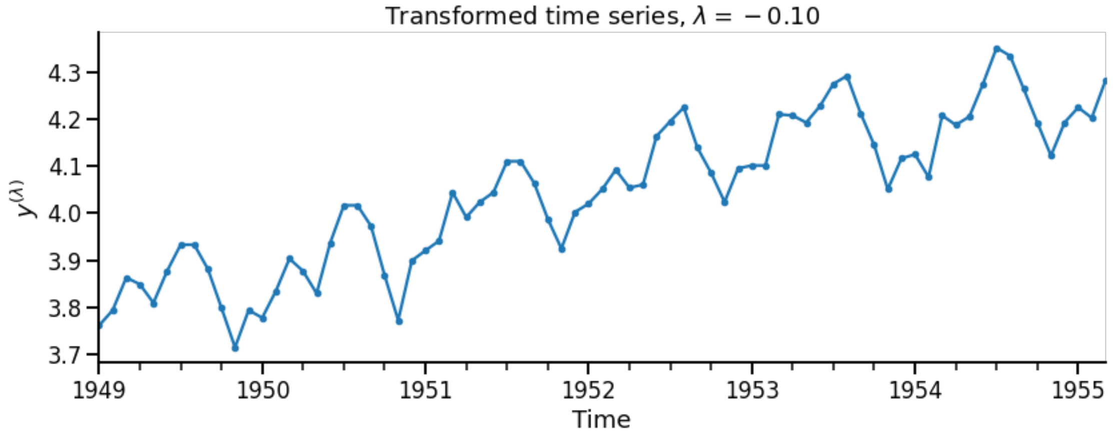


$$C_V(\lambda) = \frac{\sigma}{\mu} = \frac{\sigma[S_h(\lambda)]}{E[S_h(\lambda)]}$$

Guerrero method

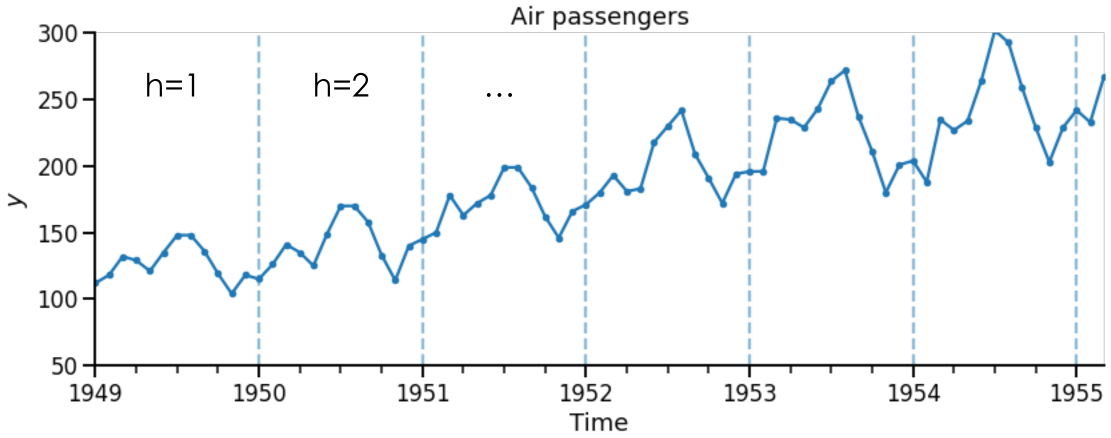
- Compute $C_V(\lambda)$ at multiple values of λ between -5 and 5.
- Pick $\lambda = \lambda_{min}$ which minimizes $C_V(\lambda)$.
- This value of λ creates a time series where $\frac{\sigma[y_t]}{E[y_t]^{1-\lambda}}$ is the "most constant" across time.
- Which implies it's the best λ to use to cause the variance of $y_t^{(\lambda)}$ to be constant.





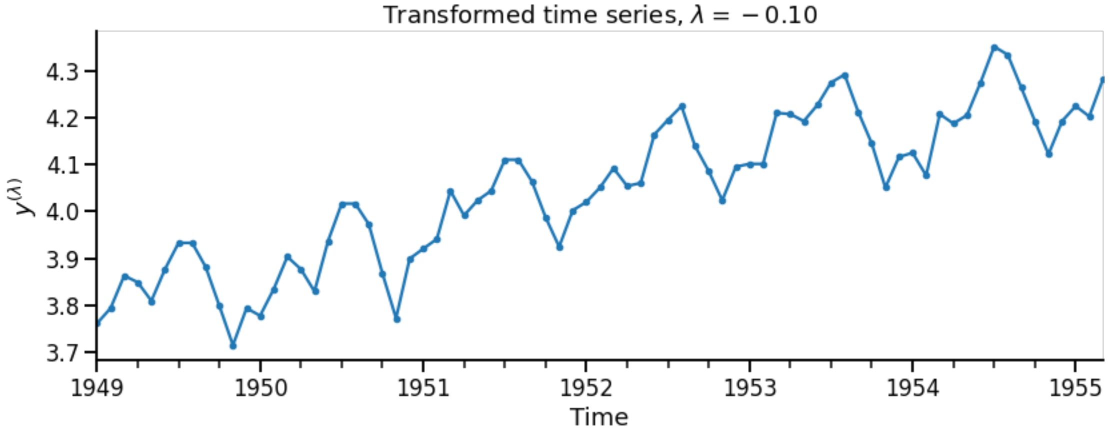
Guerrero method

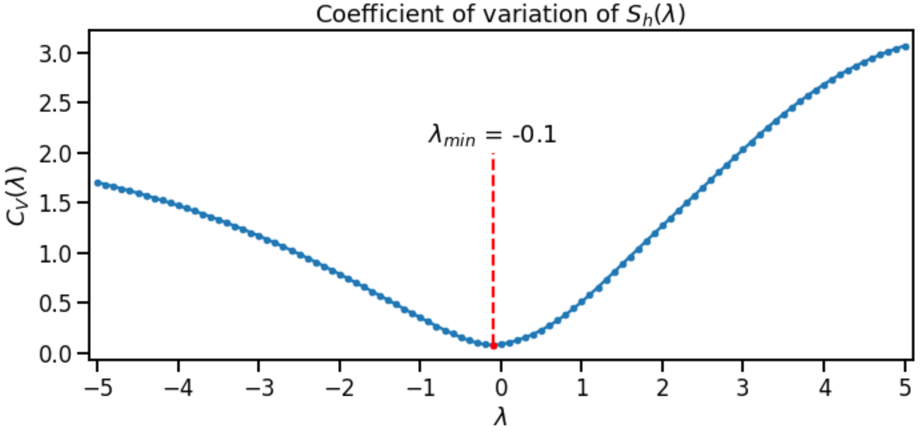
- Main parameter is the number of subseries, H, to split the original time series into.
- If the data has seasonality then split the subseries by the seasonal period (e.g., one subseries for each year if monthly data).
- If no seasonality, then split the timeseries into consecutive groups of size 2 to minimize loss of information caused by grouping.



Why use the Guerrero method?

- Does not make any assumptions about the distribution of the data.
- Directly tries to stabilize the variance across the time series.
- Therefore, more relevant for our time series tasks (i.e., forecasting and decomposition).





Box Cox implementation in sktime

BoxCoxTransformer

class **BoxCoxTransformer**(*bounds=None*, *method='mle'*, *sp=None*) [source]

Box-Cox power transform.

Box-Cox transformation is a power transformation that is used to make data more normally distributed and stabilize its variance based on the hyperparameter lambda. [1]

The BoxCoxTransformer solves for the lambda parameter used in the Box-Cox transformation given *method*, the optimization approach, and input data provided to *fit*. The use of Guerrero's method for solving for lambda requires the seasonal periodicity, *sp* be provided. [2]

Parameters:

bounds : *tuple*

Lower and upper bounds used to restrict the feasible range when solving for the value of lambda.

method : *{"pearsonr", "mle", "all", "guerrero"}, default="mle"*

The optimization approach used to determine the lambda value used in the Box-Cox transformation.

sp : *int*

Seasonal periodicity of the data in integer form. Only used if method="guerrero" is chosen. Must be an integer >= 2.

```
from sktime.transformations.series.boxcox import BoxCoxTransformer
```

```
transformer = BoxCoxTransformer(method='guerrero', sp=12)
data['y_g'] = transformer.fit_transform(data['y'])
transformer.lambda_
```

```
-0.10000000000001741
```

Summary

Forecasting and decomposition methods sometimes work better if the variance is stable across the whole time series.

A Box Cox transform can stabilize the variance, but we need to pick a good value for the parameter λ.

Guerrero method selects λ that makes the variance of $y^{(\lambda)}$ constant by minimizing the coefficient of variation.

Appendix: Guerrero method

- We denote the Box Cox Transform of a variable Y as $T(Y)$.

- We want the variance of the transformed variable to be constant:

$$Var[T(Y)] = c$$

- Taylor expand $T(Y)$ about the mean of Y, $\mathrm{E}[Y]$, to first order:

$$T(Y) \approx T(E[Y]) + T'(E[Y])(E[Y] - Y) \text{ where } T'(Y) = \frac{\partial T}{\partial Y} = Y^{\lambda-1}$$

$$\Rightarrow \; Var[T(Y)] \approx T'(E[Y])^2 Var[Y] = c$$

$$\Rightarrow \frac{Var[Y]^{\frac{1}{2}}}{E[Y]^{1-\lambda}} = \sqrt{c} = a$$