Prediction improvement of non-stationary time series analysis based on transformation

Seung-Ho Ryu[a] • Young Eun Jeon[b] • Suk-Bok Kang[b] • Jung-In Seo[c]

[a] Department of Data Science, Andong National University, Andong, Korea
[b] Department of Statistics, Yeungnam University, Gyeongsan, Korea
[c] Department of Information Statistics, Andong National University, Andong, Korea

1. Abstract

Motivation and Purpose

- A Box-Cox transformation is used to enhance the predictive performance of a non-stationary time series dataset, but it has a disadvantage of depending on a power parameter to be estimated from the dataset.
- A scaled logit transformation has no parameters to be estimated from the dataset, so it is stress-free in parameter estimation.
- This study aims to demonstrate the superiority of the scaled logit transformation, compared to the Box-Cox transformation.

Practical Example

- We demonstrate the applicability and excellence of the scaled logit transformation using the bike sharing dataset obtained from University of California, Irvine (UCI) Machine Learning Repository (https://archive.ics.uci.edu).

2. Methodology

2.1. Transformations

❑ Box-Cox Transformation (BC Trans)

$$Y_t^* = \begin{cases} \dfrac{Y_t^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log Y_t, & \lambda = 0, \end{cases}$$

where $Y_t > 0$ and λ is a power parameter.

❑ Scaled Logit Transformation (SL Trans)

$$Y_t^* = \log\left(\frac{Y_t - a}{b - Y_t}\right), a < Y_t < b,$$

where a and b are constants that must be assigned.

2.2. Traditional Time Series (TTS) Models

❑ Dynamic Harmonic Regression (DHR)

$$Y_t = \beta_0 + \sum_{i=1}^{j} \beta_{i,t} x_{i,t} + \sum_{i=1}^{K}\left[\alpha_{i,t} \sin\left(\frac{2\pi i t}{m}\right) + \gamma_{i,t} \cos\left(\frac{2\pi i t}{m}\right)\right] + \eta_t,$$

where Y_t is the observed time series data point at time t,
$\beta_{i,t}$ is the coefficient for the ith covariate $x_{i,t}$ at time t,
$\alpha_{i,t}$ and $\gamma_{i,t}$ are unobserved stochastic time variable parameters,
η_t is modeled as an autoregressive integrated moving average (ARIMA) process,
m is a seasonal frequency.

❑ ARIMA Model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d \eta_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t, \epsilon_t \sim WN(0, \sigma_\epsilon^2),$$

where p and q are the orders of the autoregressive and moving average terms,
d is the number of differences required to make a stationary time series.

❑ Seasonal-Trend decomposition using LOESS Model (STLM)

$$Y_t = A_t + S_t,$$

where A_t is the seasonally adjusted component (trend + remainder components) and predicted by regression with ARIMA error,
S_t is a seasonal component and predicted by the seasonal naive method.

❑ Bayesian Structural Time Series (BSTS)

$$\begin{aligned} Y_t &= \mu_t + \tau_t + \boldsymbol{\beta}^T \boldsymbol{x}_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \delta_t + u_t, u_t \sim N(0, \sigma_u^2), \\ \delta_{t+1} &= \delta_t + v_t, v_t \sim N(0, \sigma_v^2), \\ \tau_{t+1} &= -\sum_{s=1}^{S-1} \tau_{t+1-s} + w_t, w_t \sim N(0, \sigma_w^2), \end{aligned}$$

where μ_t is the level of the trend at time t,
δ_t is the slope of the trend at time t,
τ_t is the seasonal component.

2.3. Machine Learning (ML) Models

❑ Random Forest (RF)

RF is a ML technique that uses the bootstrap and feature bagging methods to create multiple decision trees and then collects the results of individual decision trees to make a final prediction.

❑ Extreme Gradient Boosting (XGBoost)

XGBoost is a ML technique based on a boosting algorithm that improves predictive performance by sequentially learning new models that complement errors in previous models.

❑ Light Gradient Boosting Machine (LightGBM)

LightGBM is an improved ML technique that has less memory and learning time than XGBoost, using methods such as gradient-based one-side sampling and exclusive feature bundling.

3. Analysis

3.1. Bike Sharing Dataset

- The bike sharing dataset consists of the hourly count of rental bikes between 2011 and 2012 in the bikeshare system with the corresponding weather and seasonal information.
  [DataSoruce : UCI (https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset)]
- Using the bike sharing dataset, we predict the hourly count of rental bikes.

❑ Values for transformations

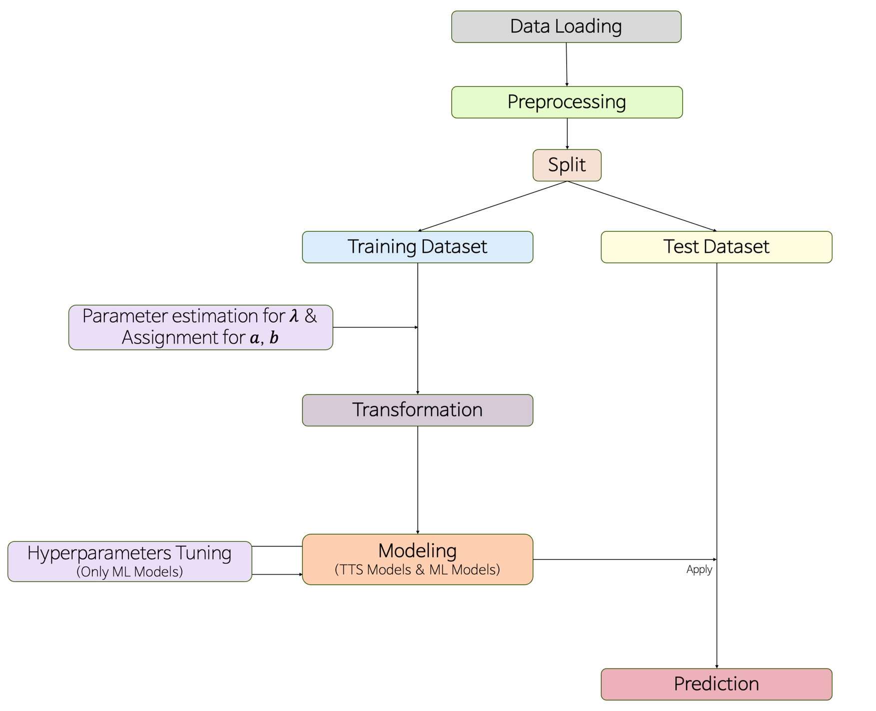
BC Trans	SL Trans
$\lambda = 0.198$	$a = -0.001$, $b = 1522.193$

- Because the number of rental bikes has a value of $\mathbf{0}$ or more, a is assigned as $\mathbf{-0.001}$, by subtracting a very small number from $\mathbf{0}$.
- Because almost all values exist within 3 standard deviations assuming that the time series dataset is normality, b is assigned as $\max(Y_{Training}) + 3\sigma = 1522.193$, where $Y_{Training}$ is the number of rental bikes of the bike sharing training dataset, σ is a standard deviation of $Y_{Training}$.

❑ Features

Feature	Type	Explaination
mnth	Categorical	Month (1 to 12)
hour	Numerical	Hour (0 to 23)
holiday	Categorical	Holiday or not
workingday	Categorical	(Holiday + Weekend) or not
weathersit	Categorical	Weather
temp	Numerical	Temperature in celsius
atemp	Numerical	Feeling temperature in celsius
hum	Numerical	Humidity

❑ Schematic Diagram



❑ Metrics

- Mean Percentage Error (MPE)
- Mean Absolute Percentage Error (MAPE)
- Symmetric Mean Absolute Percentage Error (SMAPE)

3.2. Results

❑ Parameter Estimation for TTS Models

	Original Scale			BC Trans			SL Trans		
	DHR	STLM	BSTS	DHR	STLM	BSTS	DHR	STLM	BSTS
β_{hour}	0.66	−0.21	0.00	0.01	0.00	0.15	0.01	0.00	0.11
$\beta_{month.2}$	13.32	−12.99	0.00	0.66	0.22	0.00	0.22	0.07	0.00
$\beta_{month.3}$	4.74	−13.21	0.00	−1.00	0.16	0.00	−0.75	−0.02	0.00
$\beta_{month.4}$	−5.14	−32.75	0.00	−2.44	0.12	0.00	−1.43	0.00	0.00
$\beta_{month.5}$	−12.23	−78.13	0.00	−1.82	−0.06	0.00	−1.14	−0.04	0.00
$\beta_{month.6}$	−4.33	−74.38	0.00	−0.89	0.04	0.00	−0.77	−0.01	0.00
$\beta_{month.7}$	−10.71	−62.77	0.00	−0.79	0.10	0.00	−0.64	0.01	0.00
$\beta_{month.8}$	−49.29	−73.48	0.00	−1.94	−0.04	0.00	−1.11	−0.05	0.00
$\beta_{month.9}$	−15.87	−66.34	0.00	−1.29	−0.28	0.00	−0.81	0.06	0.00
$\beta_{month.10}$	8.00	−85.85	0.00	−0.33	0.07	0.00	−0.30	−0.07	0.00
$\beta_{month.11}$	−4.40	−55.36	0.00	−0.31	0.14	0.00	−0.36	0.02	0.00
$\beta_{month.12}$	4.29	−26.02	0.00	−0.29	0.27	0.00	−0.41	0.11	0.00
$\beta_{holiday.1}$	−4.98	−10.84	0.00	−0.00	−0.01	0.00	0.01	−0.02	0.00
$\beta_{workingday.1}$	−5.74	3.54	0.00	−0.17	0.02	0.00	0.00	0.00	0.00
$\beta_{weathersit.2}$	0.41	−4.52	0.00	0.02	−0.04	0.00	0.01	−0.02	0.00
$\beta_{weathersit.3}$	−2.38	−16.96	0.00	−0.04	−0.25	−0.21	−0.02	−0.14	−0.41
β_{temp}	58.14	6.66	0.00	0.51	0.04	1.45	0.57	0.00	0.56
β_{atemp}	−6.11	5.93	0.00	0.05	0.12	0.00	−0.02	0.08	0.00
β_{hum}	−19.42	−40.02	0.00	−0.30	−0.75	−1.85	−0.25	−0.37	−1.10
α_1	15.90	–	–	0.00	–	–	0.00	–	–
α_2	−8.98	–	–	0.14	–	–	0.08	–	–
γ_1	−27.89	–	–	−0.41	–	–	−0.17	–	–
γ_2	29.93	–	–	0.50	–	–	0.24	–	–

	Original Scale		BC Trans		SL Trans	
	DHR	STLM	DHR	STLM	DHR	STLM
ϕ_1	–	1.64	0.30	1.50	0.38	1.34
ϕ_2	–	−0.91	−0.65	−0.72	0.02	−0.52
ϕ_3	–	–	–	−0.02	−0.15	−0.12
ϕ_4	–	–	–	–	−0.16	–
θ_1	0.32	−1.63	0.25	−1.49	–	−1.44
θ_2	− 0.15	0.49	0.73	0.52	–	−0.49
θ_3	–	0.38	0.25	–	–	−0.01
θ_4	–	−0.02	–	–	–	–
θ_5	–	−0.20	–	–	–	–

❑ Optimal Hyperparameters for ML Models

RF	XGBoost	LightGBM
mtry = 12	nrounds = 949	nrounds = 949
ntree = 500	eta = 0.181	eta = 0.181
nodesize = 7	max_depth = 10	max_depth = 10
	gamma = 0.547	gamma = 0.547
	min_child_weight = 11	min_child_weight = 11
	subsample = 0.944	subsample = 0.944
	colsample_bytree = 0.646	colsample_bynode = 0.646

❑ TTS Models

❑ ML Models

4. Conclusion

- The ML models have higher accuracy than the TTS models, and RF model with SL Trans has the best predictive performance among models.
- Our results strongly support the usefulness of the SL Trans and the ML models in modeling and predicting the non-stationary time series dataset.
- The SL Trans is applicable to various real non-stationary datasets such as a temperature prediction dataset.