2019 Fall WSU R working group

Time Series Analysis Basic to Bayesian Structural Time Series

By Jikhan Jeong

Contents

OLS Y_t (Dependent variable) = $\boldsymbol{Time\,Trend\,Effect_t} + \boldsymbol{Seasonal\,Effect_t} + \sum \beta \cdot X_t + \boldsymbol{Error_t}$

Serial Correlation

Time Series

1. Checking Autocorrelation: Auto Correlation Function (ACF), **Partial Auto Correlation (PAC)**
→ Determine how much lag put on AR(p)

1. Checking Cross Correlation
→ Determine how much lag put on ARDL(q)

1. Checking Unit Root
→ Determine the time series data is stationary or non-stationary

Time Series Data

Frequency Approach care for stand error of residuals

1. PAC = Determine the lag value of Y (Example, AR(1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \epsilon_t$$

2. Cross Correlation = Determine lag value of X (Example, ARDL(1,1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + \epsilon_t$$

3. Unit Root Test = Determine the stationary or not stationary

No unit root

unit root, possible spurious regression

Stationary

1) $\mathrm{E}(y_t) = \mu < \infty$
2) $var(y_t) = \sigma^2 < \infty$
3) $cov(y_t, y_{t-s}) = \gamma_s$

AR(p)
MA(q)
ARIMA
ARDL(p,q)
↓ (Multivariate)
VAR (Vector Autocorrelation) for Impulse Response

Non - Stationary

Cointegration Test

yes

VECM
(Long-term relationship)

no

LSTM
(Deep Learning)

Bayesian Approach

1. State-Space Model
2. Kalman Filter
3. Smoother
4. Spike and Slab prior
- Variable Selection
5. MCMC for posterior distribution prediction by averaging

Bayesian Structural Time Series Model

For Nowcasting,
Short-term Prediction
With High-dimensional Data

Time Series Data

Frequency Approach care for stand error of residuals

1. PAC = Determine the lag value of Y (Example, AR(1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \epsilon_t$$

2. Cross Correlation = Determine lag value of X (Example, ARDL(1,1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + \epsilon_t$$

3. Unit Root Test = Determine the stationary or not stationary

No unit root

unit root, possible spurious regression

Stationary

1) $\mathrm{E}(y_t) = \mu < \infty$
2) $var(y_t) = \sigma^2 < \infty$
3) $cov(y_t, y_{t-s}) = \gamma_s$

AR(p)
MA(q)
ARIMA
ARDL(p,q)
↓ (Multivariate)
VAR (Vector Autocorrelation) for Impulse Response

Non - Stationary

Cointegration Test

VECM
(Long-term relationship)

LSTM
(Deep Learning)

Example of Time Series Analysis Flow in the case of Non-stationary

: no code demonstration, and very quick over-view

Two endogenous variables

$$Y_1 \leftrightarrow Y_2$$

Two variables are interrelated
(Feedback Relationship)

Y_1 = JCC price
Y_2 = LNG price

Time Series Data

Frequency Approach care for stand error of residuals

1. PAC = Determine the lag value of Y (Example, AR(1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \epsilon_t$$

2. Cross Correlation = Determine lag value of X (Example, ARDL(1,3))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + \beta_3 X_{t-2} + \beta_3 X_{t-3} + \epsilon_t$$

3. Unit Root Test = Determine the stationary or not stationary

unit root, possible spurious regression

Non - Stationary

Cointegration Test

VECM
(Long-term
relationship)

Y_1 = JCC price
Y_2 = LNG price

Two endogenous variables

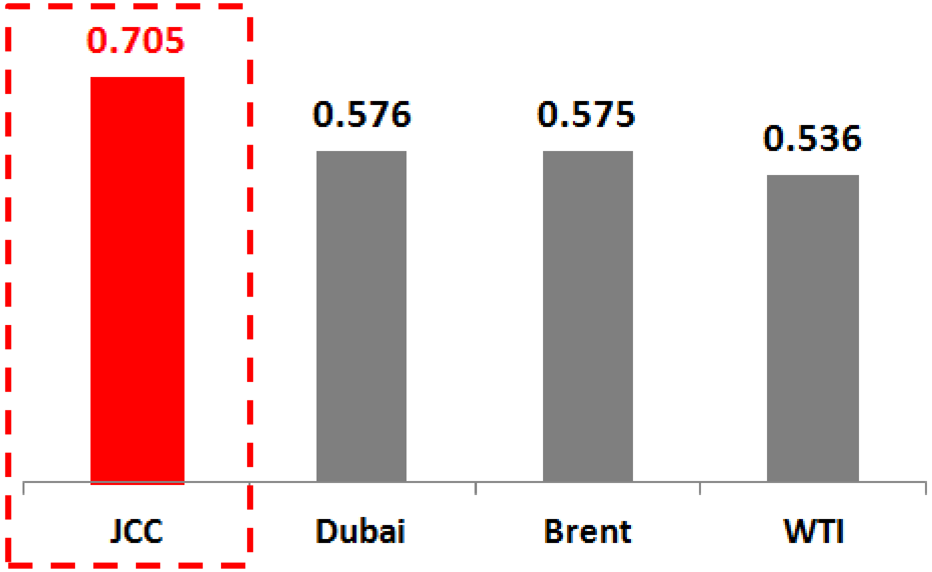
$$Y_1 \leftrightarrow Y_2$$

Two variables are interrelated

(Feedback Relationship)

Descriptive Statistics

• Correlation test between imported LNG price for power generation excluding delivery costs and oil prices



• Descriptive Statistics

Variable	Period	Mean	Max	Min
Imported LNG Price for Power Generation excluding delivery costs	2008/01 - 2016/08	15,160 Won/GJ	22,500 Won/GJ	8,551.1 Won/GJ
JCC	2008/01 - 2016/08	88 $/Bbl	135 $/Bbl	30 $/Bbl

Granger Causality

- There is a **bidirectional relationship** between JCC price and LNG price excluding delivery costs for power generation.

Dependant Variable : LNG price for power generation excluding delivery cost

Excluded	Chi Square	DF	Probability
JCC	87.408	4	0.000
ALL	87.408	4	0.000

The null hypothesis is ‘LNG price does not Granger-cause JCC’. In this case, we can reject the null that LNG price does Granger-cause JCC price.

Dependant Variable : JCC price

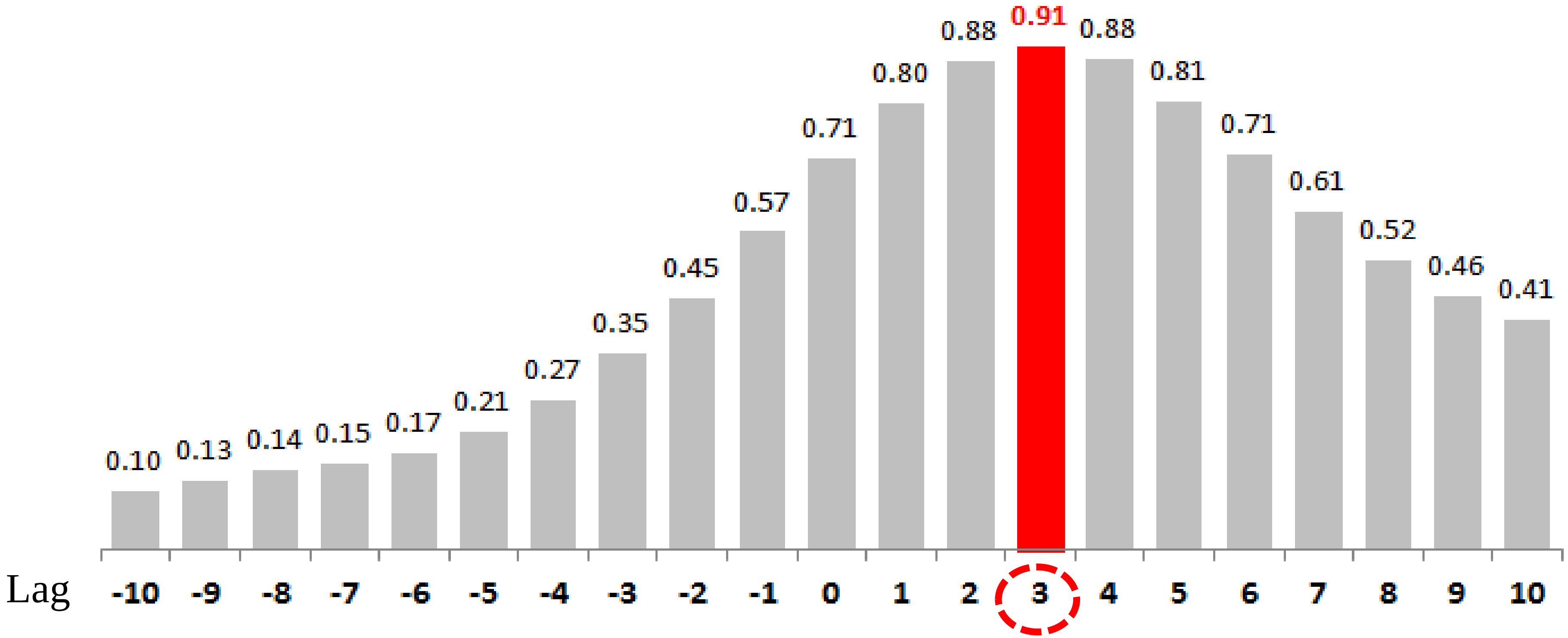
Excluded	Chi Square	DF	Probability
LNG price for power generation	13.333	4	0.010
ALL	13.333	4	0.010

The null hypothesis is ‘JCC does not Granger-cause LNG price’. In this case, we can reject the null that JCC price does Granger-cause LNG price.

Cross-correlation Analysis

• JCC price have a positive effect on future level of LNG price for power generation excluding delivery cost, reaching the highest point at three months. In this case, **JCC price are positive correlated with LNG price three months later.**

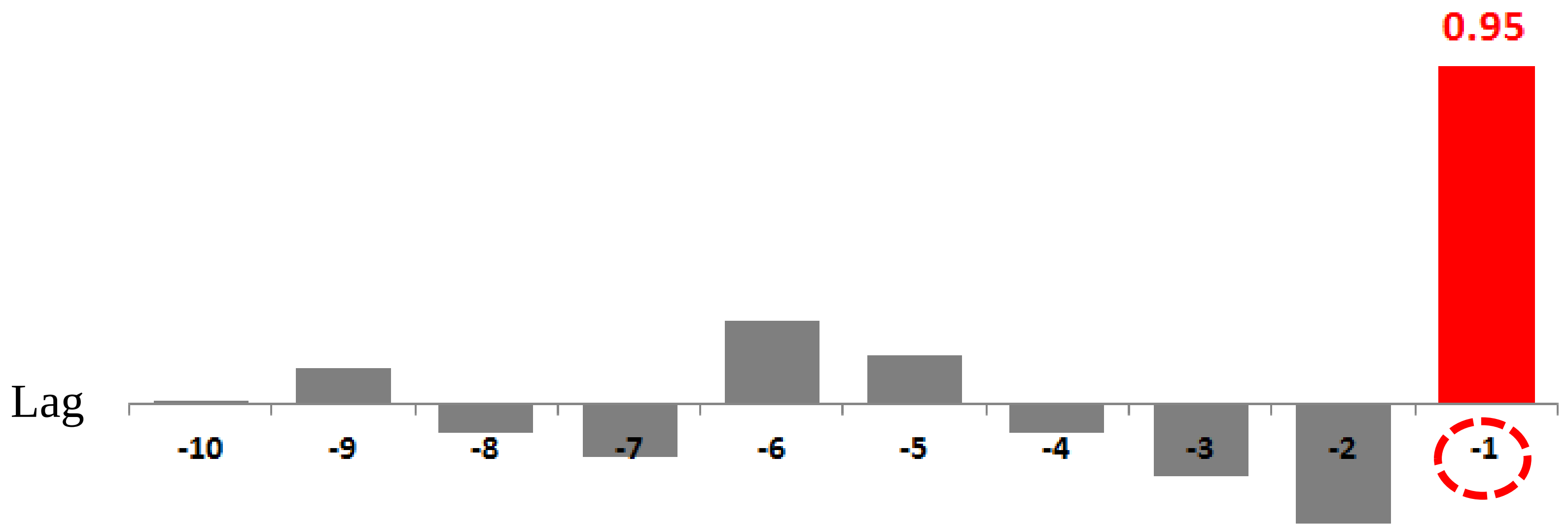
Dependant Variable : Imported LNG price for power generation excluding delivery cost
Independent Variable : JCC price



Autocorrelation Analysis

- The correlation between LNG price variable and its previous values.
- **PAC test shows that the correlation between the current value of LNG price variable and its value one month ago is 0.95**

* PAC (Partial autocorrelation) Test



Unit root test

- The Augmented Dickey-Fuller (ADF) test is one of the most commonly used unit root test
- **The ADF tests denote that LNG and JCC price variable have a unit root.**

LNG price variable : Unit root

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.264	-4.039	-3.450	-3.150

Mackinno Approximate P-value for Z(t) : 0.4541

JCC price variable : Unit root

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.687	-4.039	-3.450	-3.150

Mackinno Approximate P-value for Z(t) : 0.2415

Testing for Cointegration

- Cointegration means times series variables share an stochastic trend.
- Engle and Granger ADF test for cointegration is a two step process and its result denotes that **LNG and JCC price variables are cointegrated**.

EG-ADF test

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-4.217	-3.509	-2.890	-2.580

Davidson and Mackinno(1993) Asymptotic critical values for the cointegration test

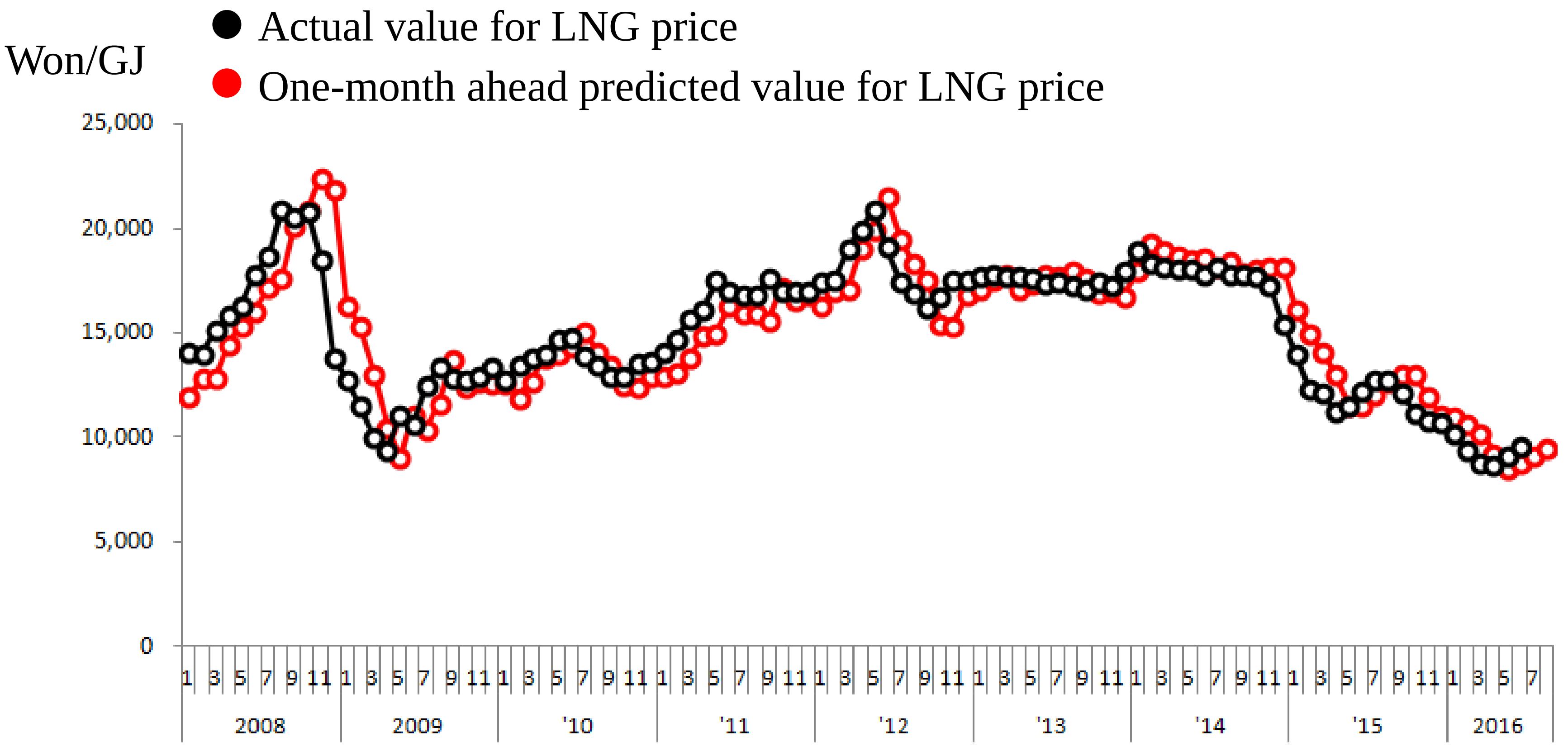
The number of variables	1% Critical Value	5% Critical Value	10% Critical Value
N = 2	-3.90	-3.34	-3.04

* The residuals in the test are not the actual error terms, but estimated values for the long run equilibrium equation of LNG price against JCC price; therefore, the critical values used EG-ADF test are not the same provided by Dickey-Fuller. Thus, we use the critical value provided by Davidson and Mackinno(1993) for the cointegration test.

No Unitroot → Reject Null Hypothesis (= No cointegration) → Both variables are cointegrated

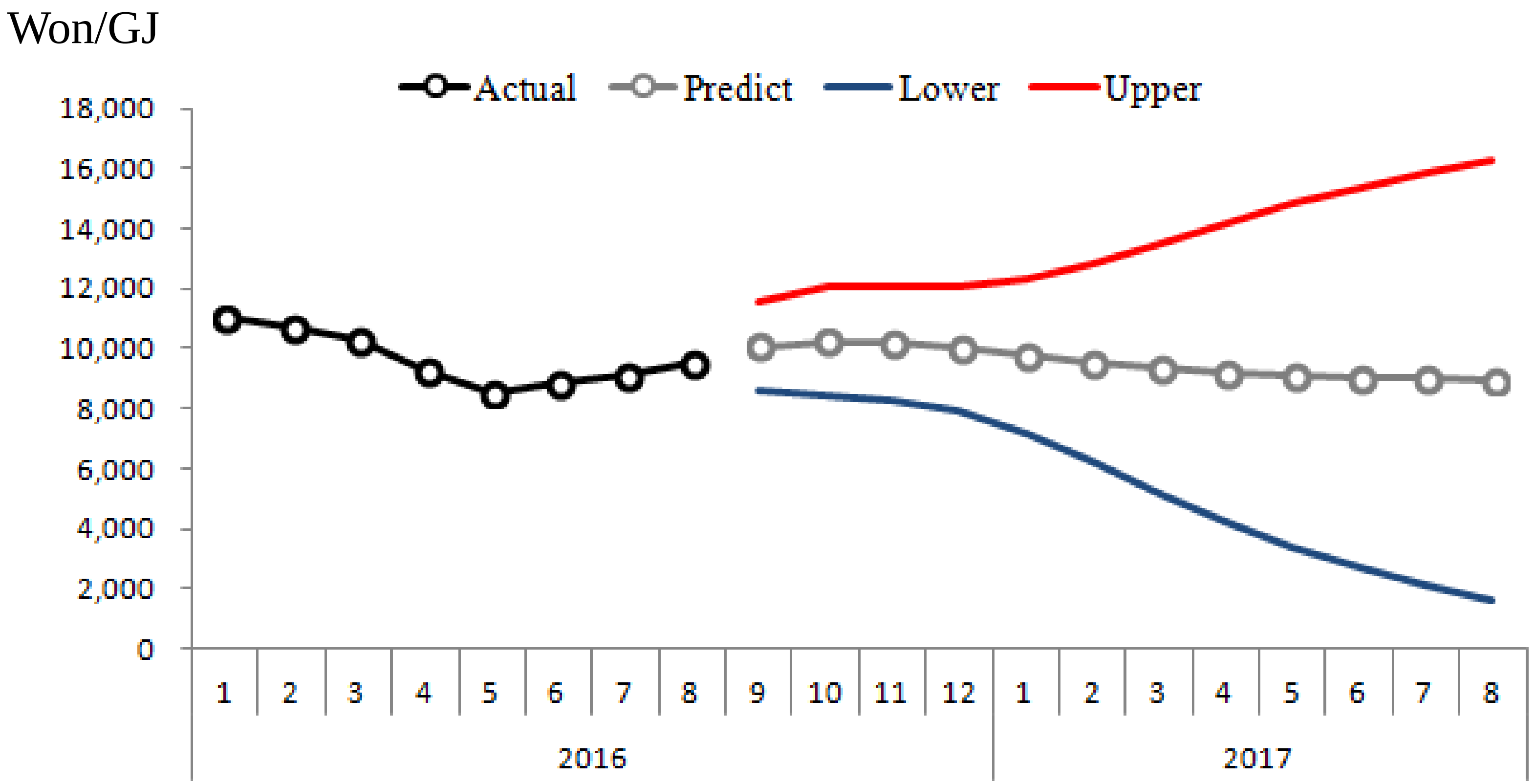
VECM (Vector Error Correction Model)

- **VEC model denotes that if the JCC price decrease 1$/Bbl, LNG price excluding delivery costs for power generation in S.Korea will decrease 114.1459 Won/GJ.**



Short-term forecasting

- **The LNG price for power generation excluding delivery cost will be likely to stay around 10,000 Won/GJ in 2017.**



Time Series Data

Frequency Approach care for stand error of residuals

1. PAC = Determine the lag value of Y (Example, AR(1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \epsilon_t$$

2. Cross Correlation = Determine lag value of X (Example, ARDL(1,1))

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 X_t + \beta_3 X_{t-1} + \epsilon_t$$

3. Unit Root Test = Determine the stationary or not stationary

No unit root

unit root, possible spurious regression

Stationary

1) $\mathrm{E}(y_t) = \mu < \infty$
2) $var(y_t) = \sigma^2 < \infty$
3) $cov(y_t, y_{t-s}) = \gamma_s$

AR(p)
MA(q)
ARIMA
ARDL(p,q)
↓ (Multivariate)
VAR (Vector Autocorrelation) for Impulse Response

Non - Stationary

Cointegration Test

VECM
(Long-term relationship)

LSTM
(Deep Learning)

Bayesian Approach

1. State-Space Model
2. Kalman Filter
3. Smoother
4. Spike and Slab prior
- Variable Selection
5. MCMC for posterior distribution prediction by averaging

Bayesian Structural Time Series Model

For Nowcasting,
Short-term Prediction
With High-dimensional Data

Time Series Data

Basic Bayesian Statistics

Posterior distribution = Prior distribution X Likelihood

$$\mathrm{P(A|B)} = \mathrm{P(A)} \times \frac{P(B|A)}{P(B)}, P(A|B)\ is\ the\ posterior\ probability$$

Bayesian Approach

1. State-Space Model
2. Kalman Filter
3. Smoother
4. Spike and Slab prior
- Variable Selection
5. MCMC for posterior distribution prediction by averaging

Bayesian Structural Time Series Model

For Nowcasting,
Short-term Prediction
With High-dimensional
Data

Bayesian Structural Time Series for Short-Term Forecasting (Nowcasting)

Predicting the Present with Bayesian Structural Time Series

Steven L. Scott
Hal Varian*

June 28, 2013

Abstract

This article describes a system for short term forecasting based on an ensemble prediction that averages over different combinations of predictors. The system combines a structural time series model for the target series with regression component capturing the contributions of contemporaneous search query data. A spike-and-slab prior on the regression coefficients induces sparsity, dramatically reducing the size of the regression problem. Our system averages over potential contributions from a very large set of models and gives easily digested reports of which coefficients are likely to be important. We illustrate with applications to initial claims for unemployment benefits and to retail sales. Although our exposition focuses on using search engine data to forecast economic time series, the underlying statistical methods can be applied to more general short term forecasting with large numbers of contemporaneous predictors.

http://people.ischool.berkeley.edu/~hal/Papers/2013/pred-present-with-bsts.pdf

Package 'bsts'

May 7, 2018

Date 2018-05-05

Title Bayesian Structural Time Series

Author Steven L. Scott <steve.the.bayesian@gmail.com>

Maintainer Steven L. Scott <steve.the.bayesian@gmail.com>

Description Time series regression using dynamic linear models fit using MCMC. See Scott and Varian (2014) <DOI:10.1504/IJMMNO.2014.059942>, among many

https://cran.r-project.org/web/packages/bsts/bsts.pdf

Bayesian Structural Time Series Basic Theory

Basic Structure

Structural Time Series: (1) + (2)

(1) The observation equation

$$y_t = Z_t^T \alpha_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2), observed\ variable, y_t,\ and\ latent\ state\ vector, \alpha_t$$

(2) The Transition equation (How latent state happens through time)

$\alpha_{t+1} = T_t \alpha_t + R_t\ \eta_t,\ \eta_t \sim N(0, \gamma^2)$, and Z_t^T, T_t, R_t are the structural parameters

Adding to latent state (1) Time trend,
(2) Seasonal effect,
(3) Features(i.e., independent variables)

Bayesian Structural Time Series Model in the example

Bayesian Structural Time Series

$y_t = \mu_t + \tau_t + \beta^T x_t + \varepsilon_t, x_t$ is the regressors

$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t,$ linear time trend

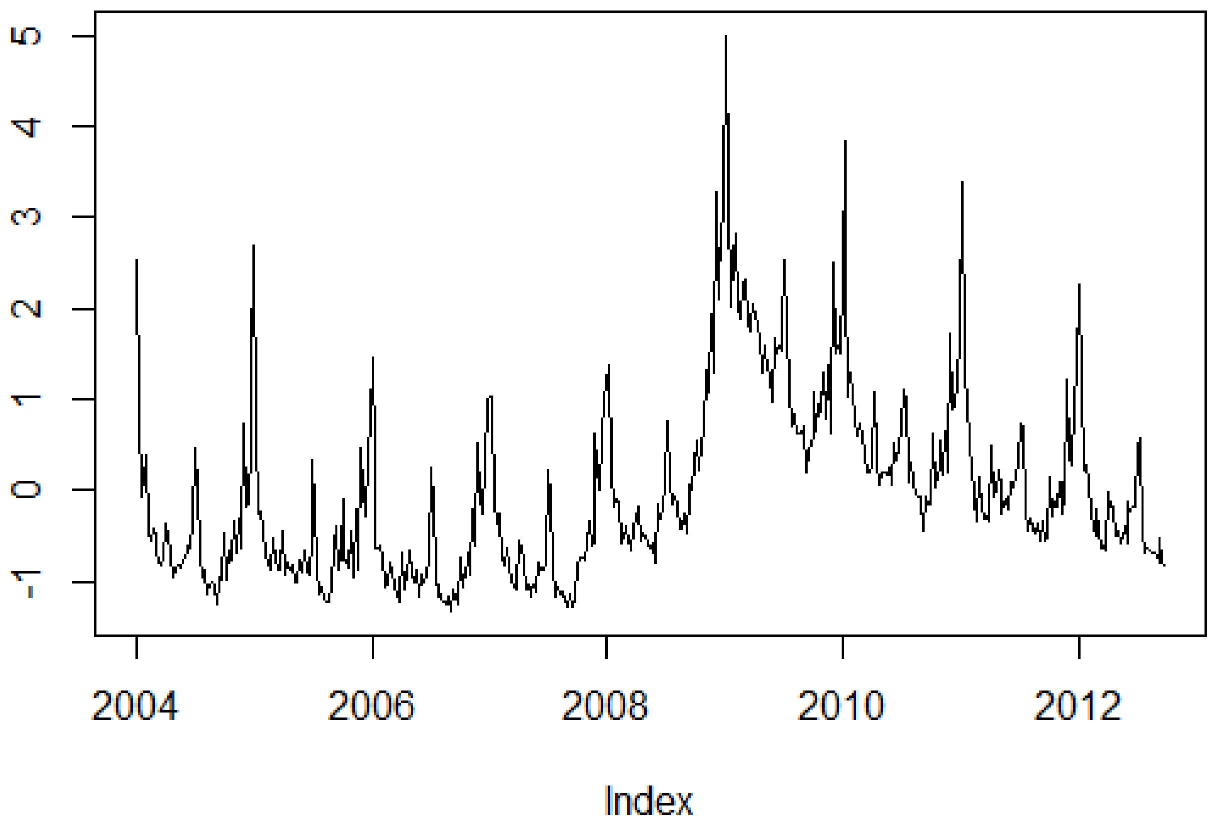
$\delta_t = \delta_{t-1} + v_t$, current slope of time trend

$\tau_t = -\sum_{s=1}^{S-1} \tau_{t-s} + w_t,$ seasonal effect

CODE practice with BSTS

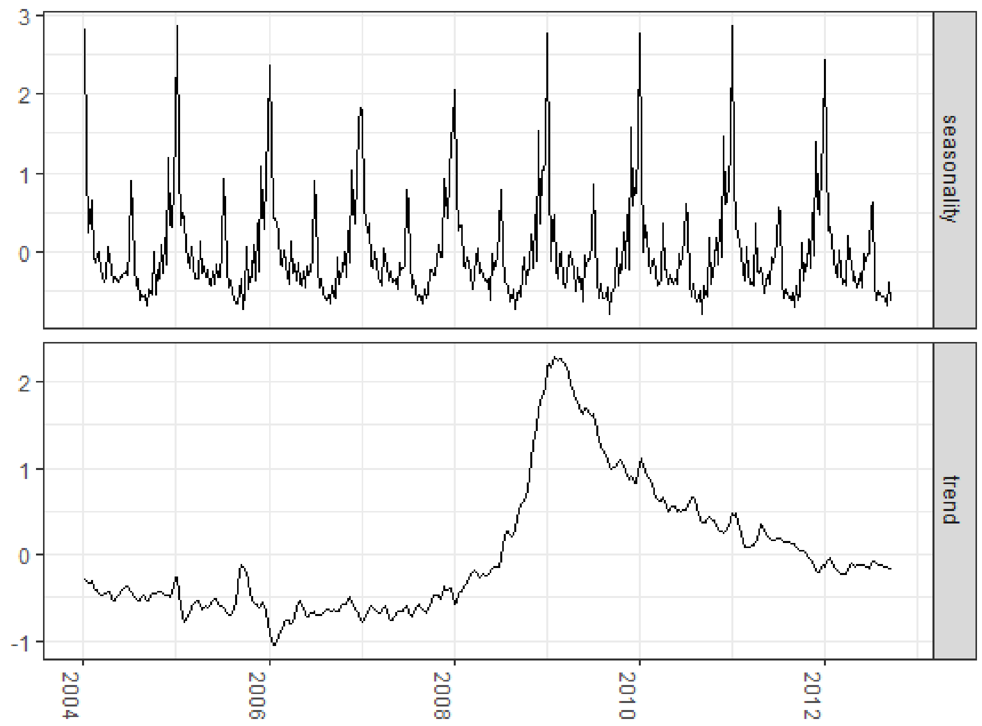
1. Install or library for **dylyr, tidyverse, bsts, BoomSpikeSlab**
2. Loading default data (iclaims)
3. Basic Model 1 (Time Trend, Seasonal Effect)
4. Model 2 (Time Trend, Seasonal Effect, 1 independent variable)
5. Model 3 (Time Trend, Seasonal Effect, 5 independent variables)
6. Compare Model1, Model2, Model3

- Loading default data (iclaims)

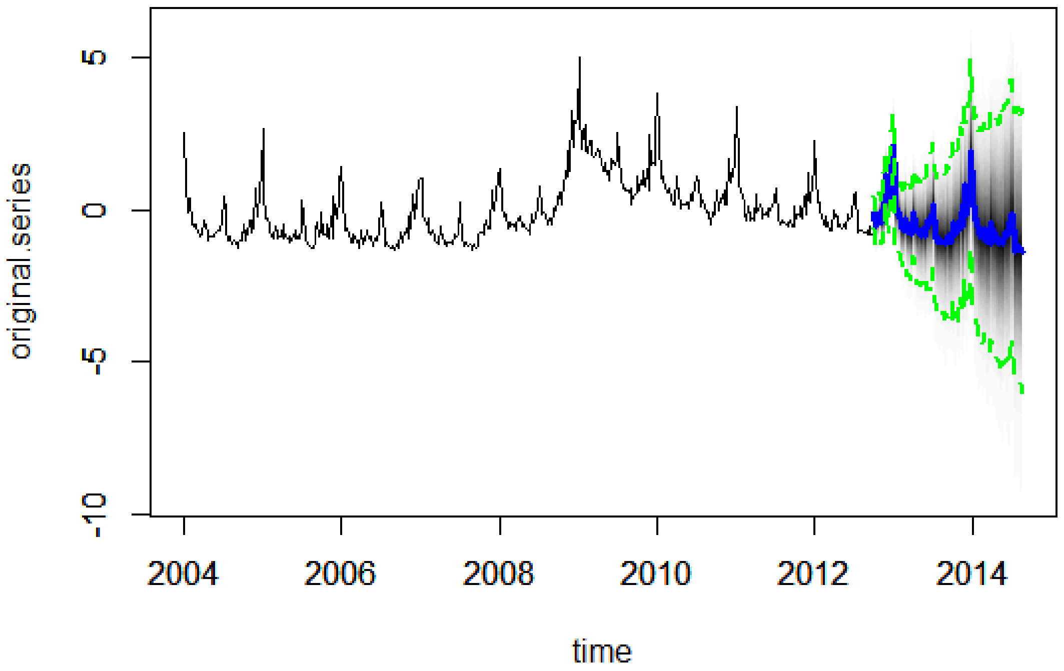


- Basic Model 1 (Time Trend, Seasonal Effect)

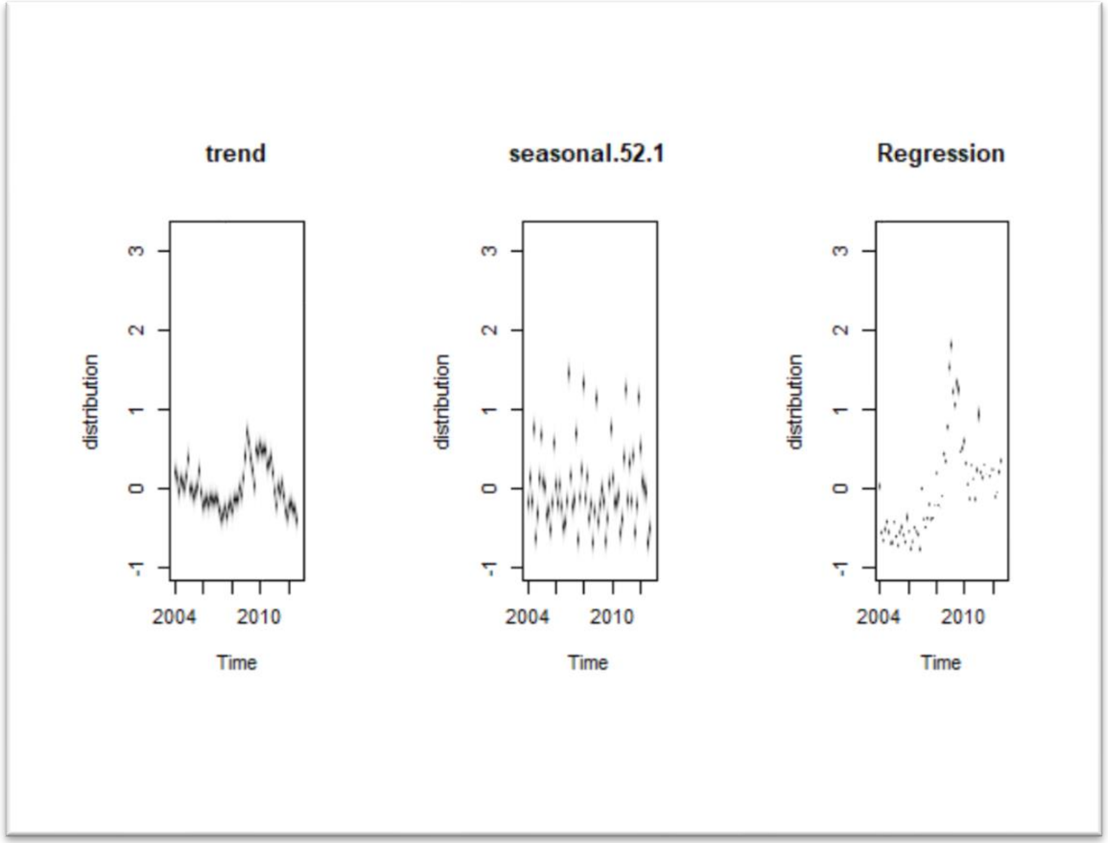
(2)______________________________Setting the basic model
2.1 making empty list()
2.2 add a time trend component
2.3 add a seasonal component
2.4 fitting a model with time and seasonal trends
2.5 prediction with fitted model

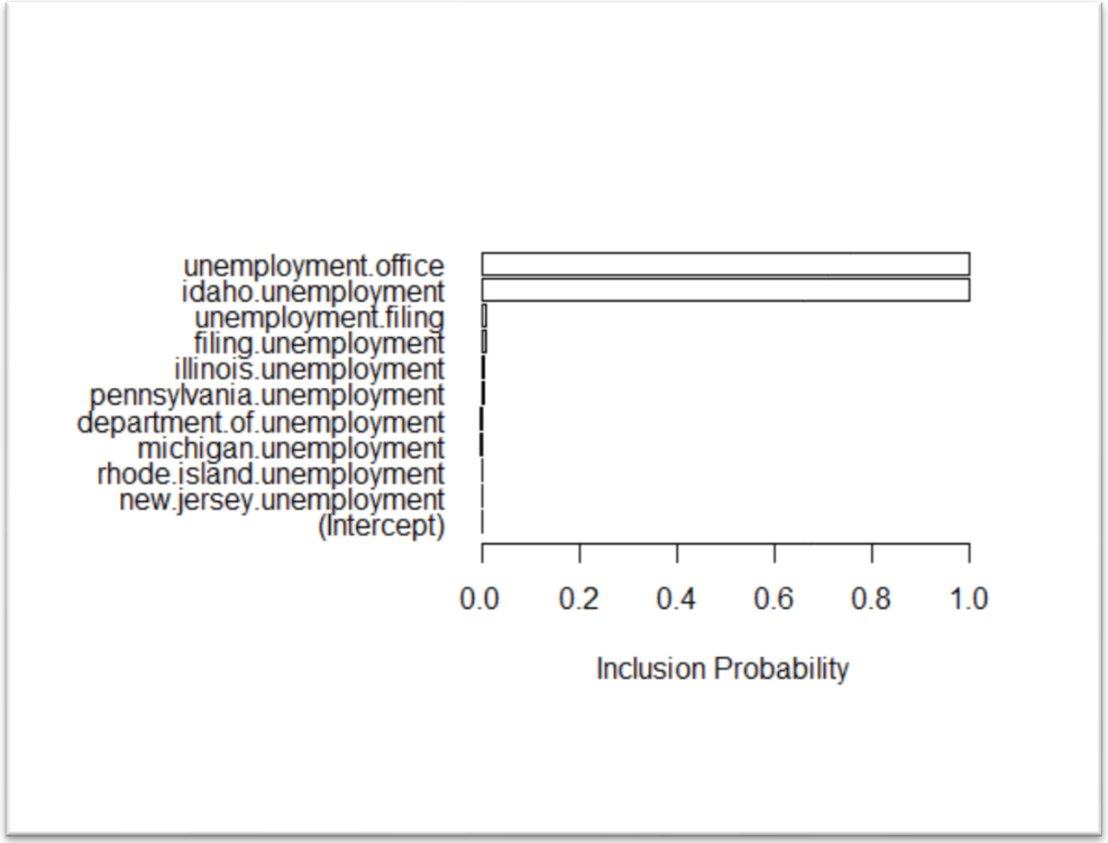


100 weeks ahead prediction

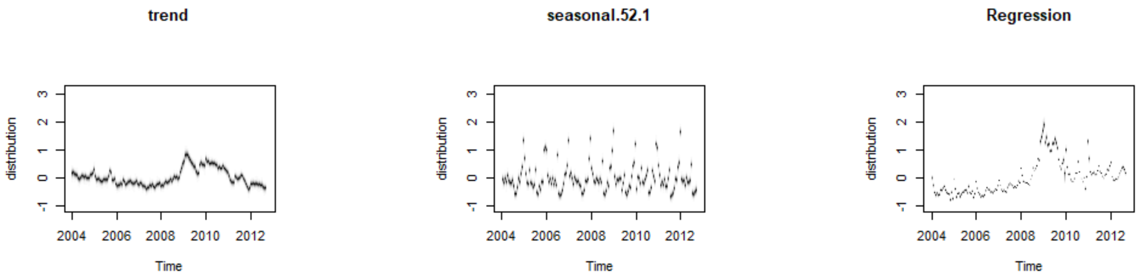


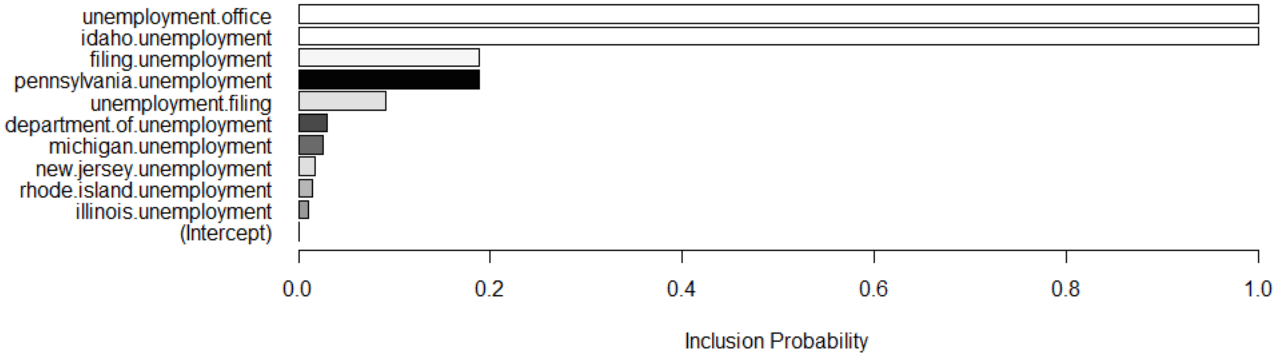
- Basic Model 3 (Time Trend, Seasonal Effect, 1 feature = default), feature selections by prior setting





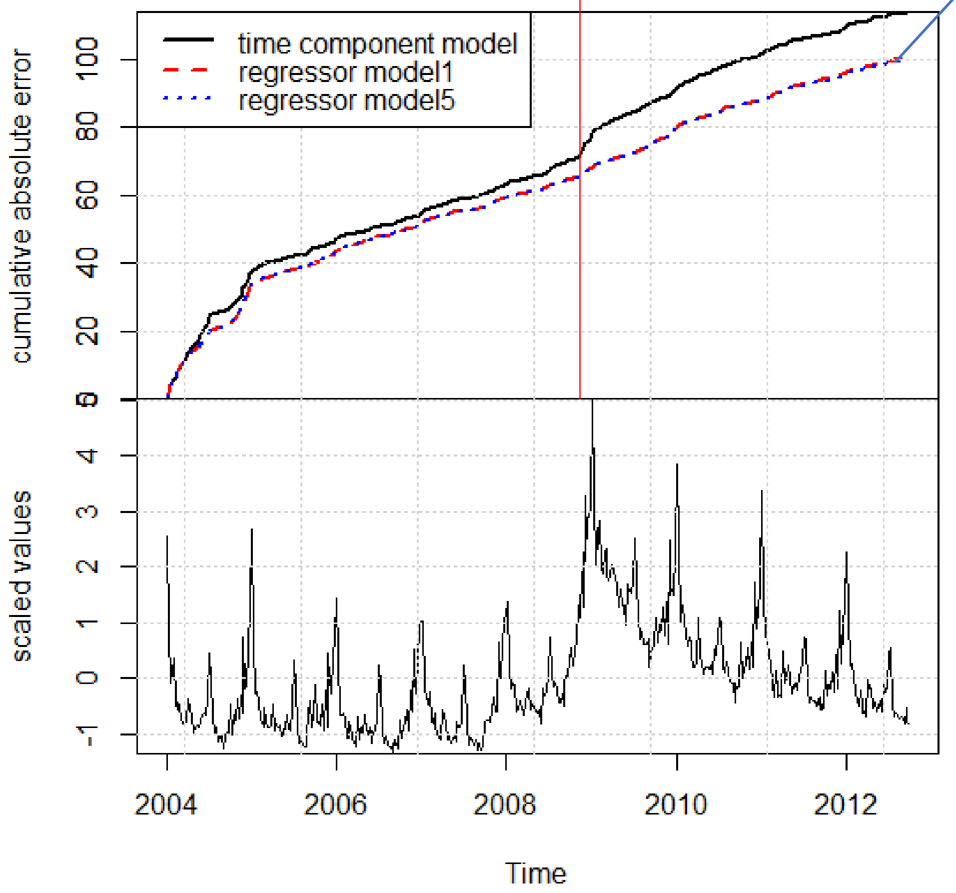
- Basic Model 4 (Time Trend, Seasonal Effect, 5 feature)





- Compare 3 models (model1 = trend, seasonal ,
  model3 = trend, seasonal, 1 feature,
  model4 = trend, seasonal, 5 features)

Turning point



Model3, model4 looks very
Similar so that added feature
Don't much work well, or redundent

Be careful,
it is 1-step ahead cumulative error

Original Data

Conclusion

- High-dimensional (p>n), you can select the variable based on prior in BSTS
- Short-term prediction is possible as a Nowcasting, i.e., predicting from the present
- You can impose your belief as a prior in the distribution

But, all the time series depends on research questions and new methods like embedding arises with era of Big-data.

Reference

- Scott, Steven, and Hal Varian. 2014. "Predicting the Present with Bayesian Structural Time Series" 5. Inderscience Publishers Ltd: 4–23.
- Kitamura, A. (Ai). (2018, September 21). Forecasting Japan's spot LNG prices using Bayesian Structural Time Series. Maritime Economics and Logistics. Retrieved from http://hdl.handle.net/2105/43636
- http://dss.princeton.edu/training/
- https://rstudio-pubs-static.s3.amazonaws.com/257314_131e2c97e7e249448ca32e555c9247c6.html
- https://www.youtube.com/user/202tylertucker
- http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html
- https://multithreaded.stitchfix.com/blog/2016/04/21/forget-arima/