

Econometrics II TA Session #13

Hiroki Kato

1 Empirical Application of Time Series Model: Nikkei 225

1.1 Background and Data

The “Nikkei225” is a stock price index published by Nihon Keizai Shimbun (hereafter, NIKKEI). NIKKEI calculates this price index based on 225 high liquid brands listed with first section of the Tokyo Stock Exchange. We use daily data of the Nikkei 225 index taken from the yahoo finance (<https://stocks.finance.yahoo.co.jp/stocks/detail/?code=998407.O>). The time length is from January 4th 2019 to January 22 2021. We have 498 observations. We read a csv data which recodes the Nikkei 225 dairy index, using the `read.csv` function in R. Since R recognize a time variable (e.g., 2021/01/22) as a character string, we need to define a time variable, using `as.Date()` function. The data structure is as follows:

```
library(tidyverse)
dt <- read.csv("data/nikkei225.csv", stringsAsFactor = FALSE)
dt$date <- as.Date(dt$date, format = "%Y/%m/%d")
head(dt)
```

##	date	open_price	high_price	low_price	close_price
## 1	2021-01-22	28580.20	28698.18	28527.16	28631.45
## 2	2021-01-21	28710.41	28846.15	28677.61	28756.86
## 3	2021-01-20	28798.74	28801.19	28402.11	28523.26
## 4	2021-01-19	28405.49	28720.91	28373.34	28633.46
## 5	2021-01-18	28238.68	28349.97	28111.54	28242.21
## 6	2021-01-15	28777.47	28820.50	28477.03	28519.18

There are five variables:

- `date`: date variable
- `open_price`: open price in day t
- `high_price`: high price in day t
- `low_price`: low price in day t
- `close_price`: closed price in day t

Mainly, we use the `date` and `close_price`.

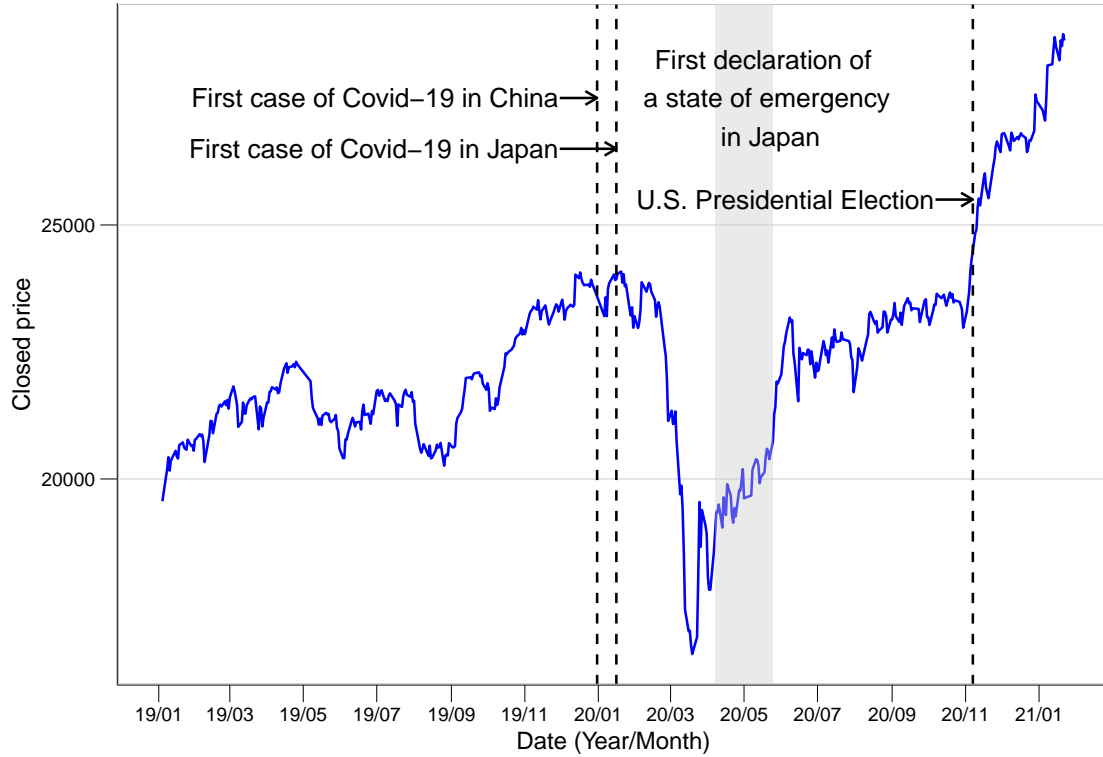


Figure 1: Time Series Data of Nikkei 225 (Closed Price)

Figure 1 shows the time series of closed price of the Nikkei225. We summarize some features as follows:

- After the COVID-19 occurred in Japan and China, the Nikkei225 has drastically decreased.
- During the first declaration of a state of emergency in Japan, the Nikkei225's performance has been a V-shaped recovery.
- The Nikkei225 has sharply increased immediately before and after the U.S. presidential election.

Some may wonder if the negative shock of COVID-19 reflects the Nikkei225. To discuss it, we need to consider following two potential concerns.

1. unlisted companies (such as small restaurant business) may suffer heavily from the negative shock of COVID-19.
2. the Nikkei225 does not represent a variation of price index of 225 brands. In principle, the Nikkei225 is a mathematical mean of stock price of 225 brands. In fact, the stock prices of top five brands which contribute to the Nikkei225 have increased at 70%. On the other hand, the stock price of other brands have decreased at 5% from the beginning of 2020 ¹.

Anyway, we test the stationarity of this time series.

¹See <https://news.yahoo.co.jp/articles/f63a4627b298857a62ac329b1ed41a88c2721bd4>.

1.2 Autoregressive of Order 1: AR(1) model

To check the stationarity of this time series, we consider the following AR(1) model:

$$X_t = \beta X_{t-1} + \epsilon_t,$$

where X_t is the closed price of Nikkei225 in day t . We assume (ϵ_t) is a white noise process, $(\epsilon_t) \sim WN(0, \sigma^2)$. that is, $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, and $Cov(\epsilon_t, \epsilon_{t+h}) = 0$ for $h \neq 0$.

To estimate the unknown parameters $\theta = (\beta, \sigma^2)$, we use the maximum likelihood method. We assume $\epsilon_t \sim N_{\mathbb{R}}(0, \sigma^2)$ for inference purposes. Then, the conditional log-likelihood is

$$M_T(X_1, \dots, X_T; \theta) = \frac{1}{T} \sum_{t=2}^T \left\{ -\log(2\pi\sigma^2) - \frac{X_t - \beta X_{t-1}}{2\sigma^2} \right\}.$$

The MLE $\hat{\theta}$ can be obtained by solving

$$\hat{\theta} = \arg \max_{\theta} M_T(X_1, \dots, X_T; \theta).$$

R provides a built-in function to estimate AR(p) model, named `ar()`. This function passes three arguments, a time-series data, estimation method, and autoregressive of order (p). To estimate AR(1) model using the maximum likelihood method, we specify the number of order, `order.max = 1`, and the method, `method = "mle"`. The R snippet is as follows:

```
ar1 <- ar(dt$close_price, method = "mle", order.max = 1)
sprintf("The estimated beta is %1.4f (s.e. = %1.4f)", ar1$ar, sqrt(ar1$asy.var.coef))

## [1] "The estimated beta is 0.9968 (s.e. = 0.0059)"

sprintf("The estimated squared sigma is %1.2f", ar1$var.pred)

## [1] "The estimated squared sigma is 74014.82"
```

1.3 Dickey-Fuller test

First, we derive the stationary condition in AR(1) model. The N -time iterated substitution of X_{t-1} yields

$$X_t = \beta^N X_{t-N} + \sum_{k=0}^N \beta^k \epsilon_{t-k}.$$

If $|\beta| < 1$, then the first term of right hand side converges to zero as $N \rightarrow \infty$. Thus, under $|\beta| < 1$, the causal stationary solution is $X_t = \sum_{k=0}^{\infty} \beta^k \epsilon_{t-k}$.

Testing for stationarity is equivalent to test for $\beta = 1$ in AR(1). The null hypothesis is that the process is **not** stationary. The alternative hypothesis is that the process is stationary, that is, $\beta < 1$. The Dickey-Fuller test provides this *one-sided* test.

To implement this test with R, we use the package called `tseries`. We use the function called `adf.test()` to carry out the Dickey-Fuller test. We need to pass two arguments in this

function. The first argument is time-series data. The second one, named `k`, is the number of order (p). In this example, we pass `k = 1`, that is, AR(1).

```
library(tseries)
df1 <- adf.test(dt$close_price, k = 1)
sprintf("The DF stats is %1.4f (p-value = %1.4f)", df1$statistic, df1$p.value)

## [1] "The DF stats is -2.7638 (p-value = 0.2550)"
```

As a result, we cannot reject the null hypothesis. Thus, the time series of closed price of Nikkei225 is not stationary when we use data from January 4th 2020 to January 22 2021.

1.4 Stationarity of Log Return of Nikkei225

We showed that the time series of closed price of Nikkei225 is not stationary. Then, is the log return of Nikkei225 stationary?

To check it, we first construct the log return of Nikkei225, using the time series of closed price. To construct log return variable, we use the `dplyr` library. First, after reading dataset `dt`, we sort dataset by date using `arrange()` function. Second, we make the lagged closed price, using `lag()` function. This function picks up the variable contained in the previous n -th row. Finally, we make the log return, using `mutate()` function which makes a new variable. As a result, the data structure is as follows:

```
dt <- dt %>%
  arrange(date) %>%
  mutate(lag_close_price = lag(close_price, n = 1)) %>%
  mutate(log_return = log(close_price) - log(lag_close_price))
head(dt[,c("date", "close_price", "lag_close_price", "log_return")])

##           date close_price lag_close_price  log_return
## 1 2019-01-04    19561.96          NA           NA
## 2 2019-01-07    20038.97    19561.96  0.024092014
## 3 2019-01-08    20204.04    20038.97  0.008203707
## 4 2019-01-09    20427.06    20204.04  0.010977908
## 5 2019-01-10    20163.80    20427.06 -0.012971575
## 6 2019-01-11    20359.70    20163.80  0.009668539
```

Figure 2 plots the time series of the log return. After the COVID-19 pandemic, the absolute value of log return is large. However, the overall time series of this index seems to be stationary.

Using `adf.test()` function in the library `tseries`, we statistically test the stationarity of time series of log return.

```
x <- dt[dt$date != as.Date("2019/01/04"), "log_return"]
df2 <- adf.test(x, k = 1)
sprintf("The DF stats is %1.4f (p-value = %1.4f)", df2$statistic, df2$p.value)
```

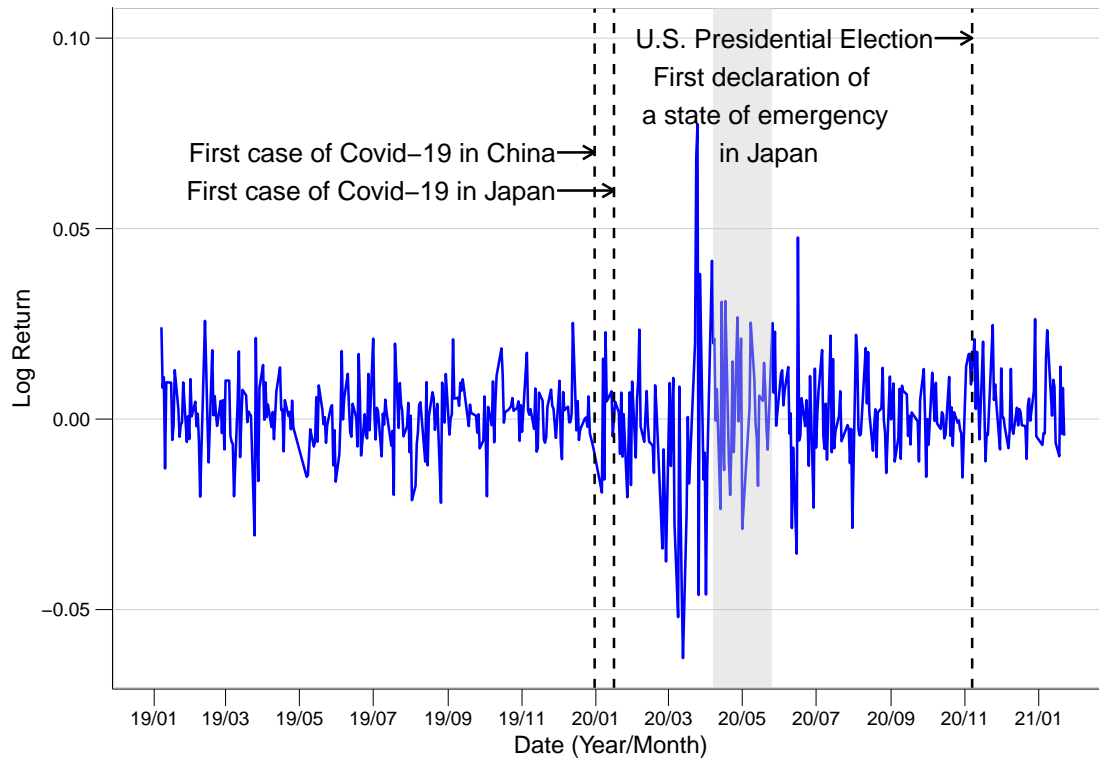


Figure 2: Time Series Data of Nikkei 225 (Log Return)

```
## [1] "The DF stats is -13.8674 (p-value = 0.0100)"
```

As a result, we can reject the null hypothesis. This implies that the time series of the log return of Nikkei225 is stationary.