Lecture 8-a1
Time Series: Introduction

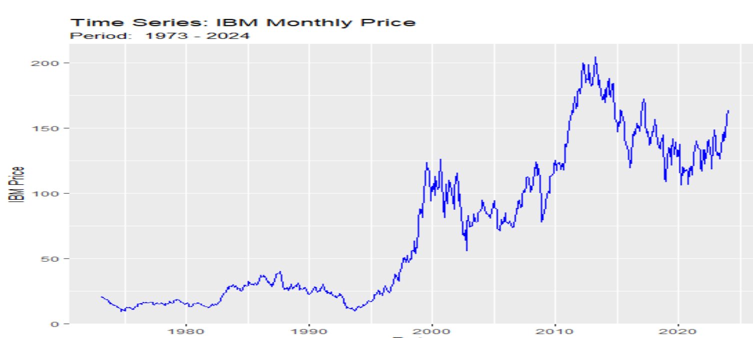
Brooks (4th edition): Chapter 6

 1

Time Series: Introduction

- A time series y_t is a process observed in sequence over time, t = 1, …., T $\Rightarrow Y_t = \{y_1, y_2, y_3, ..., y_T\}$.

Examples: IBM monthly stock prices from 1973:January till 2024:September (plot below); or USD/GBP daily exchange rates from February 15, 1923 to March 19, 1938.



Time Series: Introduction

Examples (continuation): Different ways to do the plot in R:

• Using plot.ts, creating a timeseries object in R:

```
# the function ts creates a timeseries object, start = 1973,1 (start of sample), frequency = 12(=monthly)
ts_ibm <- ts(x_ibm, start=c(1973,1), frequency=12)

plot.ts(ts_ibm,xlab="Time",ylab="IBM price", main="Time Series: IBM Stock Price")
```

• Using R package ggplot2

```
x_ibm <- SFX_da$IBM
x_date <- as.Date(SFX_da$Date, "%m/%d/%Y")
df <- data.frame(x_date, x_ibm)
ggplot(df, aes(x = x_date, y = x_ibm)) +
  geom_line(color="blue") +
  labs(x = "Date", y = "IBM Price", col = "blue", title = "Time Series: IBM Monthly Price",
     subtitle = "Period:  1973 - 2024")
```

Time Series: Introduction – Categories

• Usually, time series models are separated into two categories:

– **Univariate** ($y_t \in R$, it is a scalar)

Example: We are interested in the behavior of IBM stock prices as function of its past.

⇒ Primary model: Autoregressions (ARs).

– **Multivariate** ($y_t \in R^m$, it is a vector-valued)

Example: We are interested in the joint behavior of IBM returns, r_{IBM}, & bond yields, b_{IBM}, as function of their past

$$y_t = \begin{bmatrix} r_{IBM,t} \\ b_{IBM,t} \end{bmatrix}$$

⇒ Primary model: Vector autoregressions (VARs).

Time Series: Introduction – Dependence

• Given the sequential nature of y_t, we expect y_t & y_{t-1} to be dependent. This is the main feature of time series: **dependence**. It creates statistical problems.

• In classical statistics, we usually assume we observe several *i.i.d.* realizations of y_t. We use $\bar{y}$ to estimate the mean.

• With several independent realizations we are able to sample over the entire probability space and obtain a "good" –i.e., consistent or close to the population mean– estimator of the mean.

• But, if the samples are highly dependent, then it is likely that y_t is concentrated over a small part of the probability space. Then, the sample mean will not converge to the mean as the sample size grows.

Time Series: Introduction – Dependence

Technical note: With dependent observations, the classical results (based on LLN & CLT) are not to valid.

• We need new conditions in the DGP to make sure the sample moments (mean, variance, etc.) are good estimators population moments. The new assumptions and tools are needed: **stationarity**, **ergodicity**, CLT for martingale difference sequences (**MDS CLT**).

Roughly speaking, **stationarity** requires constant moments for y_t; **ergodicity** requires that the dependence is short-lived, eventually y_t has only a small influence on y_{t+k}, when k is relatively large.

Ergodicity describes a situation where the expectation of a random variable can be replaced by the time series expectation.

Time Series: Introduction – Dependence

An **MDS** is a discrete-time martingale with mean zero. In particular, its increments, ε_t's, are uncorrelated with any function of the available dataset at time t. To these ε_t's we will apply a CLT.

• The amount of dependence in y_t determines the 'quality' of the estimator. There are several ways to measure the dependence. The most common measure: **Covariance**.

$$\text{Cov}(y_t, y_{t+k}) = E[(y_{t_t} - \mu)(y_{t+k} - \mu)]$$

Note: When μ = 0, then $\text{Cov}(y_t, y_{t+k}) = E[y_t\, y_{t+k}]$

Time Series: Introduction – Forecasting

• In a time series model, we describe how y_t depends on past y_t's. That is, the information set is $I_t = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$

• The purpose of building a time series model: Forecasting.

• We estimate time series models to forecast out-of-sample. For example, the *l-step ahead* forecast: $\hat{y}_{T+l} = \text{E}_t[y_{t+l} \,|I_t]$.

Historical Note: In the 1970s it was found that very simple time series models out-forecasted very sophisticated (big) economic models.

This finding represented a big shock to the big multivariate models that were very popular then. It forced a re-evaluation of these big models.

Time Series: Introduction – White Noise

• In general, we assume the error term, ε_t, is uncorrelated with everything, with mean 0 and constant variance, σ^2. We call a process like this a **white noise (WN) process**.

• We denote a WN process as

$$\varepsilon_t \sim \text{WN}(0, \sigma^2)$$

• White noise is the basic building block of all time series. It can be written as simple function of a WN(0, 1) process:

$$z_t = \sigma\, u_t, \qquad u_t \sim i.i.d.\ \text{WN}(0, 1) \qquad \Rightarrow z_t \sim \text{WN}(0, \sigma^2)$$

• The z_t's are random shocks, with no dependence over time, representing unpredictable events. It represents a model of news.

Time Series: Introduction – Conditionality

• We make a key distinction: *Conditional* & *Unconditional* moments. In time series we model the conditional mean as a function of its past, for example in an AR(1) process, we have:

$$y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t.$$

Then, the **conditional mean** forecast at time t, conditioning on information at time I_{t-1}, is:

$$\text{E}_t[y_t\,|\,I_{t-1}] = \text{E}_t[y_t] = \alpha + \beta\, y_{t-1}$$

Notice that the **unconditional mean**, μ, is given by:

$$\text{E}[y_t] = \alpha + \beta\, \text{E}[y_{t-1}] = \frac{\alpha}{1-\beta} = \mu = \text{constant} \qquad (\beta \neq 1)$$

The conditional mean is time varying; the unconditional mean is not!

Key distinction: Conditional vs. Unconditional moments.

Time Series: Introduction – AR and MA models

• Two popular models for $E_t[y_t | I_t]$:

– An **autoregressive** (**AR**) **process** models $E_t[y_t | I_{t-1}]$ with lagged dependent variables:

$$E_t[y_t|I_t] = f(y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p})$$

Example: AR(1) process, $y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t$.

– A **moving average** (**MA**) **process** models $E_t[y_t|I_t]$ with lagged errors, ε_t:

$$E_t[y_t | I_t] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \dots, \varepsilon_{t-q})$$

Example: MA(1) process, $y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t$

• There is a third model, **ARMA**, that combines lagged dependent variables and lagged errors.

Time Series: Introduction – Forecasting (again)

• We want to select an appropriate time series model to forecast y_t. In this class, we will use linear models, with choices: AR(p), MA(q) or ARMA(p, q).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine p, q.

(2) Estimate the model.

(3) Test the model.

(4) Forecast.

• In this lecture, we go over the statistical theory (stationarity, ergodicity), the main models (AR, MA & ARMA) and tools that will help us describe and identify a proper model.

CLM Revisited: Time Series Implications

• With autocorrelated data, we get dependent observations. For example, with autocorrelated errors:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t,$$

the independence assumption is violated. The LLN and the CLT cannot be easily applied in this context. We need new tools.

• We introduce the concepts of **stationarity** and **ergodicity**. The ergodic theorem will give us a counterpart to the LLN.

To get asymptotic distributions, we also need a CLT for dependent variables, using new technical concepts: mixing and stationarity. Or we can rely on a new CLT: The *martingale difference sequence CLT*.

• We will not cover these technical points in detail.

Time Series – Stationarity

• Consider the joint probability distribution of the collection of RVs:

$$F(y_{t_1}, y_{t_2}, \ldots, y_{t_T}) = F(Y_{t_1} \leq y_{t_1}, Y_{t_2} \leq y_{t_2}, \ldots, Y_{t_T} \leq y_{t_T})$$

To do statistical analysis with dependent observations, we need extra assumptions. We need some form of invariance on the structure of the time series.

If the distribution F is changing with every observation, estimation and inference become very difficult.

• Stationarity is an invariant property: The statistical characteristics of the time series do not change over time.

• There different definitions of stationarity, they differ in how strong is the invariance of the distribution over time.

Time Series – Stationarity

• We say that a process is **stationary** of

1st order if $F(y_{t_1}) = F(y_{t_{1+k}})$ for any t_1, k

2nd order if $F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}})$ for any t_1, t_2, k

Nth-order if $F(y_{t_1}, ..., y_{t_T}) = F(y_{t_{1+k}}, ..., y_{t_{T+k}})$ for any t_1, ..., t_T, k

• *Nth-order* stationarity is a strong assumption (& difficult to verify in practice). *2nd order* (weak) stationarity is weaker. **Weak stationarity** only considers means & covariances (easier to verify in practice).

• Moments describe a distribution. We calculate moments as usual:

$$\mathrm{E}[Y_t] = \mu$$

$$\mathrm{Var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2]$$

$$\mathrm{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \gamma(t_1 - t_2)$$

Time Series – Stationarity & Autocovariances

• $\mathrm{Cov}(Y_{t_1}, Y_{t_2}) = \gamma(t_1 - t_2)$ is called the **auto-covariance function**. It measures how y_t, measured at time t_1, and y_t, measured at time t_2, covary.

Notes: $\gamma(t_1 - t_2)$ is a function of $k = t_1 - t_2$
$\gamma(0)$ is the variance.

• The autocovariance function is symmetric. That is,

$$\gamma(t_1 - t_2) = \mathrm{Cov}(Y_{t_1}, Y_{t_2}) = \mathrm{Cov}(Y_{t_2}, Y_{t_1}) = \gamma(t_2 - t_1)$$

$$\Rightarrow \gamma(k) = \gamma(-k)$$

• Autocovariances are unit dependent. We have different values if we calculate the autocovariance for IBM returns in % or in decimal terms.

Remark: The autocovariance measures the (linear) dependence between two Y_t's separated by k periods.

Time Series – Stationarity & Autocorrelations

• From the autocovariances, we derive the **autocorrelations**:

$$\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2}) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1}\sigma_{t_2}} = \frac{\gamma(t_1 - t_2)}{\gamma(0)}$$

the last step takes assumes: $\sigma_{t_1} = \sigma_{t_2} = \sqrt{\gamma(0)}$

• $\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2})$ is called the **auto-correlation function** (**ACF**), –think of it as a function of $k = t_2 - t_1$. The ACF is also symmetric.

• Unlike autocovoriances, autocorrelations are not unit dependent. It is easier to compare dependencies across different time series.

• Stationarity requires all these moments to be independent of time. If the moments are time dependent, we say the series is **non-stationary**.

Time Series – Stationarity & Constant Moments

• For a strictly stationary process (constant moments), we need:

$$\mu_t = \mu$$
$$\sigma_t = \sigma$$

because $F(y_{t_1}) = F(y_{t_{1+k}}) \Rightarrow \quad \mu_{t_1} = \mu_{t_{1+k}} = \mu$, $\quad \sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$

Then,

$$F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}}) \Rightarrow \text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$

$$\Rightarrow \rho(t_1, t_2) = \rho(t_{1+k}, t_{2+k})$$

Let $t_1 = t - k$ & $t_2 = t$

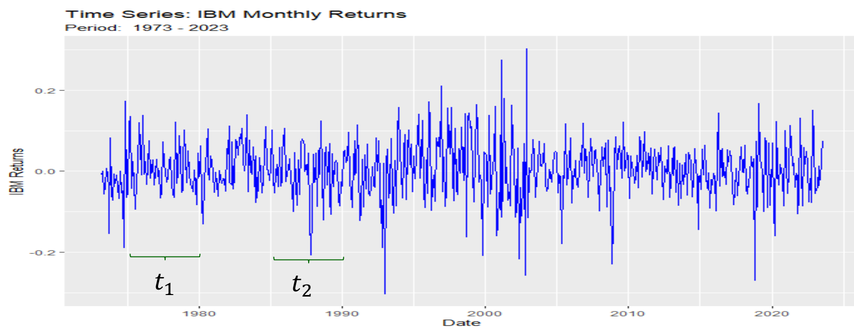
$$\Rightarrow \rho(t_1, t_2) = \rho(t-k, t) = \rho(t, t-k) = \rho(k) = \rho_k$$

The correlation between any two RVs depends on the time difference. Given the symmetry, we have $\rho(k) = \rho(-k)$.

Time Series – Stationarity & Constant Moments

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

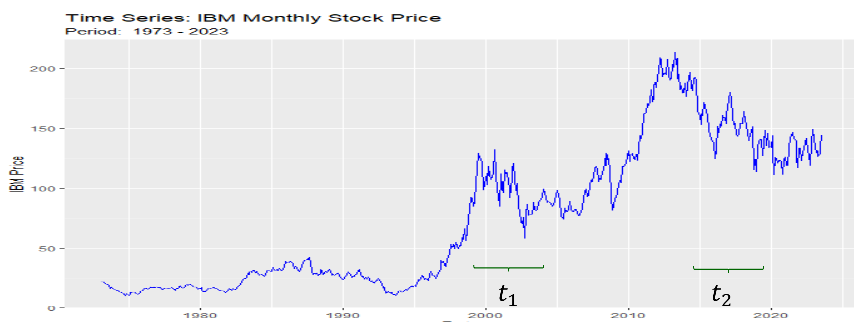
$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Stationarity & Constant Moments

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Weak Stationary

- A **Covariance stationary** process (or *2nd -order weakly stationary*) has:
  - constant mean, μ
  - constant variance, σ^2
  - covariance depends on time difference, k, between two RVs, $\gamma(k)$

That is, Z_t is covariance stationary if:

$$E(Z_t) = \text{constant} = \mu$$
$$\text{Var}(Z_t) = \text{constant} = \sigma^2$$
$$\text{Cov}(Z_{t_1}, Z_{t_2}) = \gamma(k = t_1 - t_2)$$

Remark: Covariance stationarity is only concerned with the covariance of a process, only the mean, variance and covariance are time-invariant.

Time Series – Stationarity: Example

Example: Assume y_t follows an AR(1) process:

$$y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

- **Mean**

Taking expectations on both side:

$$E[y_t] = \phi\, E[y_{t-1}] + E[\varepsilon_t]$$
$$\mu = \phi\, \mu + 0$$
$$E[y_t] = \mu = 0 \qquad \text{(assuming } \phi \neq 1)$$

- **Variance**

Applying the variance on both side:

$$\text{Var}[y_t] = \gamma(0) = \phi^2\, \text{Var}[y_{t-1}] + \text{Var}[\varepsilon_t]$$
$$\gamma(0) = \phi^2\, \gamma(0) + \sigma^2$$
$$\gamma(0) = \frac{\sigma^2}{1-\phi^2} \qquad \text{(assuming } |\phi| < 1)$$

Time Series – Stationarity: Example

Example (continuation): $y_t = \phi\, y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\begin{aligned}\gamma(1) &= \text{Cov}[y_t, y_{t-1}] = \text{E}[y_t\, y_{t-1}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-1}] \\ &= \phi\, \text{E}[y_{t-1}\, y_{t-1}] + \text{E}[\varepsilon_t\, y_{t-1}] \\ &= \phi\, \text{E}[y_{t-1}{}^2] \\ &= \phi\, \text{Var}[y_{t-1}{}^2] \\ &= \phi\, \gamma(0)\end{aligned}$$

$$\begin{aligned}\gamma(2) &= \text{Cov}[y_t, y_{t-2}] = \text{E}[y_t\, y_{t-2}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-2}] \\ &= \phi\, \text{E}[y_{t-1}\, y_{t-2}] \\ &= \phi\, \text{Cov}[y_t, y_{t-1}] \\ &= \phi\, \gamma(1) \\ &= \phi^2\, \gamma(0)\end{aligned}$$

$\vdots$

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

Time Series – Stationarity: Example

Example (continuation): $y_t = \phi\, y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

$\Rightarrow$ If $|\phi| < 1$, y_t process is covariance stationary: mean, variance, and covariance are constant.

Remark: To establish stationarity, we need to impose conditions on the AR parameters. (Conditions are not needed for MA processes.)

Note: From the autocovariance function, we derive ACF:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi^k\, \gamma(0)}{\gamma(0)} = \phi^k$$

If $|\phi| < 1$, autocovariance function & ACF show exponential decay.

Time Series – Non-Stationarity: Example

Example: Assume y_t follows a Random Walk with drift process:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Doing backward substitution:

$$\begin{aligned} y_t &= \mu + (\mu + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= 2 * \mu + y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &= 2 * \mu + (\mu + y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \varepsilon_{t-1} \\ &= 3 * \mu + y_{t-3} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} \end{aligned}$$

$$\Rightarrow y_t = \mu\, t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$$

- **Mean** & **Variance**

$$\text{E}[y_t] = \mu\, t + y_0$$

$$\text{Var}[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2\, t$$

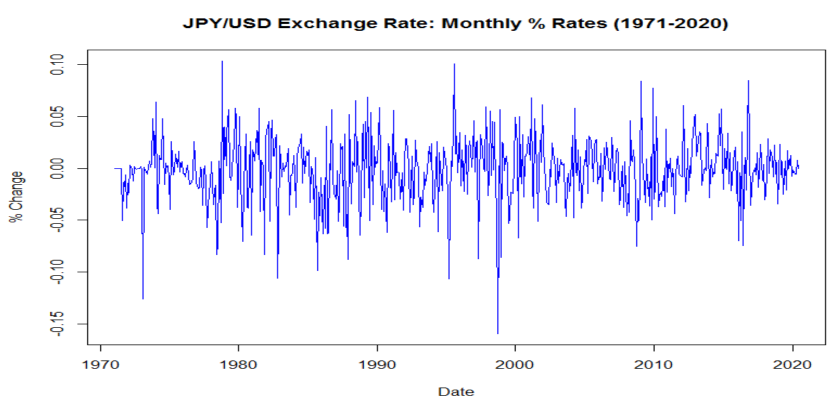
$\Rightarrow$ the process y_t is non-stationary: moments are time dependent.

Stationary Series: Examples

Examples: Assume $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

$y_t = 0.08 + \varepsilon_t + 0.4\, \varepsilon_{t-1}$ - MA(1) process

$y_t = 0.13\, y_{t-1} + \varepsilon_t$ - AR(1) process

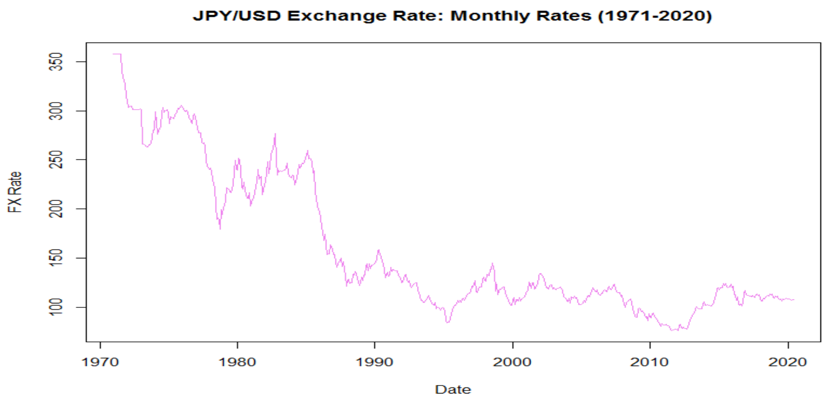


Non-Stationary Series: Examples

Examples: Assume $\varepsilon_t \sim WN(0, \sigma^2)$.

$y_t = \mu\, t + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \varepsilon_t$ - AR(2) with deterministic trend

$y_t = \mu + y_{t-1} + \varepsilon_t$ - Random Walk with drift



Time Series – Stationarity: Remarks

- Main characteristic of time series: Observations are **dependent**.

- If we have non-stationary series (say, mean or variance are changing with each observation), it is not possible to make inferences.

- Stationarity is an invariant property: the statistical characteristics of the time series do not vary over time.

- If IBM is weak stationary, then, the returns of IBM may change month to month or year to year, but the average return and the variance in two equal-length time intervals will be more or less the same.

Time Series – Stationarity (Again)

- In the long run, say 100-200 years, the stationarity assumption may not be realistic. After all, technological change has affected the return of IBM over the long run. But, in the short-run, stationarity seems likely to hold.

- In general, time series analysis is done under the stationarity assumption.