STA457S 2023 Summer

Anton S.

- **Professor:** Esam Mahdi, `e.mahdi@mail.utoronto.ca`
- **Lectures:** MW 6-9 BA1160

Contents

1. Introduction

Time series can be defined as a collection of random variables indexed according to the order they are obtained in time. We can consider time series as a sequence of random variables

$$x_1, x_2, \ldots, x_t, \ldots$$

where x_t is obtained at t-th time point. In this course, the indexing variable t will typically be discrete and not continuous. I.e. $t \in \mathbb{N}$ or $t \in \mathbb{Z}$. A *time series* is a series of observed values (x_t), we call the unrealized model a *process* in this course.

Definition 1.0.1. A series is *stationary* if it remains around a mean value over time.

Examples: Daily temperature, stock prices, generally measurements

Box-Jenkins Methodology

1. **Identification:** Examine graphs and identify patterns and dependency in an observed time series. We look for: *trend, periodic trend, outliers, irregular change*
2. **Estimation:** Select a suitable fitted model for predicting future values.
3. **Diagnostic checking:** Goodness of fit tests and residual scores to estimate adequacy of the model, determine unaccounted for patterns.
4. **Forecasting:** Use model to forecast the future values.

We say forecasting instead of prediction to indicate foretelling closely into the future.

Financial Time Series

We motivate a lot of this course with financial data, so we define terminology for financial time series.

Definition 1.0.2. The *net return* from the holding period $t-1$ to t is

$$R_t = \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1$$

i.e. relative percent increase of (x_k) from $t-1$ to t.

Definition 1.0.3. The *simple gross return* from the holding period $t-1$ to t is

$$\frac{x_t}{x_{t-1}} = 1 + R_t$$

Definition 1.0.4. The *gross return over the most recent k periods* is defined as

$$1 + R_t(k) = \frac{x_t}{x_{t-k}} = \prod_{i=0}^{i=k} \frac{x_{t-i}}{x_{t-i-1}} = (1+R_t)\ldots(1+R_{t-k})$$

Definition 1.0.5. The *log returns* or *continuously compounded returns* are denoted r_t and defined as

$$r_t = \log(1+R_t) = \log(x_t) - \log(x_{t-1})$$

Returns are scale-free but not unitless since they depend on t.

Definition 1.0.6. The *volatility* is the conditional standard deviation of underlying asset return.

In most financial time series data, the scale of the volatility appears to be the same. Highly volatile periods tend to be clustered together.

We may decompose a financial time series as

$$x_t = \underbrace{T_t}_{\text{trend}} + \underbrace{s_t}_{\text{season}} + \underbrace{c_t}_{\text{cycle}} + \underbrace{I_t}_{\text{irregularity}}$$

If these components are corelated, use a multiplicative decomposition $x_t = T_t s_t c_t I_t$. If only some are corelated, use a mixed model, i.e. $x_t = s_t T_t + c_t + I_t$.

Time Series Models

Definition 1.0.7 (Moving average)**.** The k-th (odd) moving average of a time series (x_t) is defined as the sum of the k values of the time series around x_t. For example, the third moving average series for (x_t) is

$$y_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

If k is even, we reindex and define the time of the moving average to be at the middle of the times we evaluate. For example the 4-th moving average of (x_t) is

$$y_t = \frac{1}{4}(x_{t-2} + x_{t-1} + x_{t+1} + x_{t+2})$$

Moving averages allow us to 'smooth' a time series by reducing the noise while maintaining the trend in the series.

Definition 1.0.8 (White noise)**.** A *white noise process* is a collection of uncorrelated and identically distributed random variables (w_t), each with 0 mean and finite variance σ_w^2 for every t. If the white noise follows a normal distribution, i.e.

$$w_t \sim N(0, \sigma_w^2)$$

then it is *Gaussian white noise*. In the Gaussian case, independent and uncorrelated are the same, so w_t are i.i.d.

Definition 1.0.9 (Random walk)**.** A *random walk with drift* (x_t) is a series

$$x_t = \delta + x_{t-1} + w_t$$

where $w_t \sim \text{wn}(0, \sigma^2)$. For $t \geqslant 1$, δ is the *drift*. When $\delta = 0$, the series is simply a random walk:

$$x_t = x_{t-1} + w_t$$

The series is the same as in the previous time step plus a white noise shock. Therefore we may write

$$x_t = \delta t + \sum_{j=1}^{t} w_j, \quad t \geqslant 1$$

If $\delta \neq 0$, the series is not stationary.

Definition 1.0.10 (Signal in noise)**.** Many realistic models for generating time series assume an underlying sinusoidal signal:

$$x_t = A\sin(\omega t + \phi) + \omega_t$$

As a general note, the goal of time series analysis is to apply a series of transformations in order to reduce the remaining model to a white noise series. Through these transformations we address trends in the series, aiming to be left with only a noise series.

2. Characteristics of Time Series

A complete description of time series is provided by the joint distribution function.

Definition 2.0.1. The *mean function* is defined as

$$\mu_t = E(x_t) = \int_{-\infty}^{\infty} x f_t(x)dx$$

μ_t is the expectation of the process at the given t, f_t is probability density of x_t.

Definition 2.0.2. The *autocovariance function* is defined as the second moment product

$$\gamma_x(s,t) = \text{Cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

for all s, t. Note $\gamma_x(t,t) = \text{Var}(x_t)$.

Covariance measures the 'linear relationship' of random variables (it is an inner product on the space). The following examples can be computed with the bilinearity properties of covariance.

Example 1. Consider white noise $w_t \sim \text{wn}(0, \sigma^2)$. Then we have

$$\gamma_w(s,t) = \text{Cov}(w_s, w_t) = \begin{cases} \sigma^2, & s = t \\ 0, & s \neq t \end{cases}$$

Example 2. Consider moving average $v_t = \frac{1}{3}(w_{t+1} + w_t + w_{t-1})$ with $w_t \sim \text{wn}(0, \sigma^2)$. Then we can verify that

$$\gamma_v(s,t) = \begin{cases} \frac{1}{3}\sigma^2, & s = t \\ \frac{2}{9}\sigma^2, & |s-t| = 1 \\ \frac{1}{9}\sigma^2, & |s-t| = 2 \\ 0, & |s-t| > 2 \end{cases}$$

Note: Prof said this is a great exam question!

Example 3. For a random walk without drift, $x_t = \sum_{j=1}^{t} w_j$ and $w_t \sim \text{wn}(0, \sigma^2)$, we have

$$\gamma_x(s,t) = \min\{s,t\}\,\sigma^2$$

since the w_t are uncorrelated random variables. Note $\text{Var}(x_t) = t\sigma^2$.

Definition 2.0.3. The **autocorrelation function** is defined as

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

The autocorrelation function gives a profile of the linear correlation of the series at time t. Cauchy Schwarz implies $|\gamma(s,t)|^2 \leqslant \gamma(s,s)\gamma(t,t)$.

Definition 2.0.4. For multivariate time series we have the **cross-variance** function

$$\gamma_{xy}(s,t) = \mathrm{Cov}(x_s, y_t)$$

and **cross-correlation** function

$$\rho_{xy}(s,t) = \frac{\sqrt{\gamma_{xy}(s,t)}}{\sqrt{\gamma_x(s,s)}\sqrt{\gamma_y(t,t)}}$$

This can be extended to time series with arbitrary components.

Stationary Models

Definition 2.0.5. A **stationary process** x_t has constant mean and variance for all t.

Stationarity is defined uniquely, so there is only one way for a series to be stationary. It is preferred that estimators of parameters do not changed over time. In many cases, stationary data can be approximated with stationary ARMA models which we discuss later. They also avoid the problem of *spurious regression*.

Definition 2.0.6. A series x_t is **strong stationary** if for any $t_1, t_2, \ldots, t_n \in \mathbb{Z}$ where $n \geqslant 1$ and any scalar shift $h \in \mathbb{Z}$, the joint distribution of both series is the same:

$$P(x_{t_1} \leqslant c_1, \cdots, x_{t_n} \leqslant c_n) = P(x_{t_1+h} \leqslant c_1, \cdots, x_{t_n+h} \leqslant c_n)$$

We never actually know the joint distribution, but this definition allows us to make some theoretical observations about time series. The above implies

1. $p(x_t \leqslant c) = p(x_{t+h} \leqslant c)$
2. $\mu_t = \mu_s$ for all s, t
3. $\gamma(s,t) = \gamma(s+h, t+h)$

It cannot be checked whether any observed time series is strong stationary. This motivates *weak stationary*.

Definition 2.0.7. A process is **time invariant** if it does not depend on time.

Definition 2.0.8. A time series is **weak stationary invariant, covariance stationary, second-order stationary** if

1. μ_t is constant
2. $\gamma(s,t) = \mathrm{Cov}(x_s, x_t)$ depends on s, t only by the difference $|s-t|$: $\gamma(t+h, t) = \gamma(h, 0)$.

Proposition 1. A strong stationary series is weakly stationary. The converse is not true.

Definition 2.0.9. The **autocovariance function of a stationary time series** will be written as

$$\gamma(h,0) = \gamma(h) = \mathrm{Cov}(x_{t+h}, x_t)$$

Note $\gamma(h) = \gamma(-h)$.

Definition 2.0.10. The **autocorrelation function of a stationary time series** will be written as

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)}\sqrt{\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}$$

Definition 2.0.11. The stochastic process w_t is a **strong white noise process** with mean zero and variance σ_w^2 and written $w_t \sim \mathrm{wn}(0,\sigma_w^2)$ if and only if it is i.i.d. with zero mean and covariance

$$\gamma_w(h) = E(w_t w_{t+h}) = \begin{cases} \sigma_w^2, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

A weak stationary *Gaussian* white noise process is strongly stationary, due to uncorrelated implying independent in this case.

Example 4. Consider moving average $v_t = \frac{1}{3}(w_{t+1} + w_t + w_{t-1})$ with $w_t \sim \mathrm{wn}(0,\sigma^2)$. It is stationary since $\mu_{v,t} = 0$.

$$\gamma_v(h) = \begin{cases} \frac{1}{3}\sigma^2, & h = 0 \\ \frac{2}{9}\sigma^2, & h = 1 \\ \frac{1}{9}\sigma^2, & h = 2 \\ 0, & h > 2 \end{cases} \qquad \rho_v(h) = \begin{cases} 1 & h = 0 \\ \frac{2}{3} & h = 1 \\ \frac{1}{3} & h = 2 \\ 0, & h > 2 \end{cases}$$

Example 5. $x_t = \varepsilon_t$ where $\varepsilon_t \sim \mathrm{i.i.d}(0,1)$ is weakly stationary.

Example 6. $x_t = t + \varepsilon_t$ where $\varepsilon_t \sim \mathrm{i.i.d}(0,1)$ is not weakly stationary since μ_t depends on t.

Example 7. Suppose $X_t = A\sin(t+B)$ where $A \sim r.v.(0,1)$, $B \sim U([-\pi,\pi])$. This process is stationary.

$$\begin{aligned} E(X_t) &= E(A\sin(t+B)) = E(A)E(\sin(t+B)) = 0 \\ \gamma(h) &= \frac{1}{2}\cos(h) \end{aligned}$$

$\gamma(h)$ can be verified by integrating.

Transforming Nonstationary Series

The random walk process $x_t = \delta t + \sum_{j=1}^{t} w_j$ is not stationary if it has drift, since $E(x_t) = \delta t$ depends on time. Suppose $\delta = 0$ so the mean function is constant. In this case

$$\gamma(h) = \mathrm{Cov}(x_t, x_{t+h}) = t\sigma^2 \text{ and } \rho(h) = \frac{\mathrm{Cov}(x_t, x_{t+h})}{\sqrt{\mathrm{Var}(x_t)\mathrm{Var}(x_{t+h})}} = \frac{1}{\sqrt{1+h/t}}$$

For large t and h much smaller than t, get $\gamma(h)$ is very close to 1. We can eliminate the stationarity in a random walk process by taking the difference of the x_t:

$$\nabla x_t = x_t - x_{t-1} = \varepsilon_t \sim \mathrm{wn}(0,\sigma_w^2)$$

In the presence of d unit rots, we apply d differences to x_t:

$$\nabla^d x_t = (1-B)^d x_t = \varepsilon_t$$

Where B is the backwards shift $Bx_t = x_{t-1}$. For example, consider $x_t = a + bt + ct^2$. Then we may take second order differences:

$$z_t = \nabla^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = 2c$$

for $t \geqslant 3$. The R function `diff(x, lag, differences)` can be used for this. The series ∇x_t can be used to transform the time series into stationarity.

Definition 2.0.12. A series is **jointly stationary** if they are each stationary and

$$\gamma_{xy}(h) = \text{Cov}(x_{t+h}, y_t) = E(x_{t+h} - \mu_x)(y_t - \mu_y)$$

is only a function of the lag. The **cross correlation function** of two jointly stationary series is

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}$$

We again have $-1 \leqslant \rho_{xy}(h) \leqslant 1$.

Example 8. Consider two series $x_t = w_t + w_{t-1}$ and $y_t = w_t - w_{t-1}$. We find the cross correlation

Definition 2.0.13. A **linear process** x_t is defined to be a linear combination

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_j, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

We may verify $\gamma_x(h) = \sum_{j=-\infty}^{\infty} \psi_{t+h}\psi_t$. Only need $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ for process to have finite variance. Note that the moving average is an example of a linear process.

If a time series is stationary, we may estimate the mean with $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$. In this case,

$$\text{Var}(\bar{x}) = \frac{\sigma_x^2}{n}\left(1 + \sum_{h=1}^{n-1}(1 - h/n)\rho(h)\right)$$

Estimators

Definition 2.0.14. The **sample autocovariance** is defined as

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$$

The sum is restricted since x_{t+h} is not available for $t + h > n$. This estimator is preferred than the one dividing by $n - h$ since it is a non-negative definite function. The **sample autocorrelation** is defined as

$$\hat{\rho}(0) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

This allows us to test whether the autocorrelation is statistically significant at some lags: for n sufficiently large, approximately we have $\hat{\rho}(h) \sim N(0, \frac{1}{n})$. I.e. the estimator is normally distributed with

$$\mu_{\hat{\rho}(h)} = 0 \text{ and } \sigma_{\hat{\rho}(h)} = \frac{1}{\sqrt{n}}$$

We can test $H_0 : \rho(h) = 0$, $H_a : \rho(h) \neq 0$. For $\alpha = 0.05$, have $|\hat{\rho}(h)| \geqslant \frac{2}{\sqrt{n}}$.

- The ACF **cuts off at lag** h if there no spikes at lags $> h$ in the ACF plot.
- The ACF **dies down** if it decreases in a steady fashion.
- If ACF dies down quickly, then the data is stationary. If it dies down very slowly, it is not stationary.

Vector valued time series

Same as regular time series, except

$$\mathbf{x}_t = (x_{t,1}, \dots, x_{t,p}) \in \mathbb{R}^p$$

The transpose is denoted $\mathbf{x}_t$. The mean is $\mu_t = E(x_t) = (\mu_{t,1}, \dots, \mu_{t,p})$. If the process is stationary, $E(x_t) = \mu$, and has autocovariance matrix

$$\Gamma(h) = E(x_{t+h} - \mu)(x_t - \mu)'$$

with cross covariance functions $\gamma_{ij}(h) = E(x_{t+h,i} - \mu_i)(x_{t,j} - \mu_j)$. Note $\gamma_{ij}(h) = \gamma_{ji}(-h)$.

Definition 2.0.15. The sample autocovariance matrix

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \overline{x})(x_t - \overline{x})'$$

where $\overline{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$. The symmetry property holds: $\hat{\Gamma}(h) = \hat{\Gamma}(-h)'$.