Analysis of single time series: Part I

Daniele Bianchi[1]

whitesphd.com

[1]School of Economics and Finance
Queen Mary, University of London

Summary

Most data in financial markets occur sequentially through time. Interest rates, asset returns, and foreign exchange rates are all examples of time series. This week we introduce time-series econometrics and focuses primarily on linear models.

The analysis of time-series data begins by defining two key concepts in the analysis of time series: stationarity and ergodicity. We then turn our attention to Autoregressive Moving Average (ARMA) models and cover the structure of these models, stationarity conditions, model selection and inference.

Contents

Stationarity and ergodicity

Stochastic processes

Let introduce first a building block of time-series analysis: stochastic processes.

A **stochastic process** is an arbitrary sequence of random data and is denoted $\{y_t\}$, where $\{\cdot\}$ is used to indicate that the ys form a sequence.

The simplest non-trivial stochastic process specifies that $y_t \overset{iid}{\sim} D$ for some distribution D, for example Normal.

Another simple stochastic process is the random walk,

$$y_t = y_{t-1} + \epsilon_t,$$

where ϵ_t is an iid process.

Stationarity

Strictly speaking, **stationarity** is a probabilistically meaningful measure of **regularity**.

This regularity can be exploited to estimate unknown parameters and characterize the dependence between observations across time.

Stationarity exists in two forms: **strict stationarity** and **covariance** (also known as weak) **stationarity**.

Definition (Strict stationarity)

A stochastic process $\{y_t\}$ is strictly stationary if the joint distribution $\{y_t, y_{t+1}, \ldots, y_{t+k}\}$ only depends on h and not t.

Simply put, strict stationarity requires that the *joint* distribution of a stocastic process does not depend on time and so the only factor affecting the relationship between two observations is the gap between them.

Stationarity

Definition (Covariance stationarity)

A stochastic process $\{y_t\}$ is covariance starionary if

$$\begin{aligned} E\left[y_t\right] &= \mu \qquad \text{for} \qquad t = 1, 2, \ldots \\ V\left[y_t\right] &= \sigma^2 < \infty \qquad \text{for} \qquad t = 1, 2, \\ E\left[\left(y_t - \mu\right)\left(y_{t-s} - \mu\right)\right] &= \gamma_s \qquad \text{for} \qquad t = 1, 2, \ldots, t-1 \end{aligned}$$

Simply put, covariance stationarity requires that both the unconditional mean and variance are finite and do not change over time.

Note that covariance stationarity only applies to *unconditional moments* and not conditional moments, and so a covariance process may have a varying conditional mean, i.e., be predictable.

Ergodicity

In essence, if an ergodic stochastic process is sampled at two points far apart in time, these samples will be independent.

The theoretical definition of ergoditicy is quite complex, but there is a more practical definition of ergodicity.

Definition (Ergodicity)

If $\{y_t\}$ is ergodic and its r^{th} moment μ_r is finite, then $T^{-1}\sum_{t=1}^{T} y_t^r \xrightarrow{r} \mu_r$.

In practice, the ergodic theorem states that averages will converge to their expectation provided the expectation exists.

White noise

The third important building block in time-series analysis is white noise.

White noise generalizes i.i.d. noise and allows for dependence in a series as long as three conditions are satisfied: the mean is zero, observations are not autocorrelated and the series has finite second moments.

Definition (White noise)

A process $\{\epsilon_t\}$ is known as white noise if

$$
\begin{aligned}
E\left[\epsilon_t\right] &= 0 \qquad \text{for} \qquad t = 1, 2, \ldots \\
V\left[\epsilon_t\right] &= \sigma^2 < \infty \qquad \text{for} \qquad t = 1, 2, \ldots \\
E\left[\epsilon_t \epsilon_{t-j}\right] &= Cov\left(\epsilon_t, \epsilon_{t-j}\right) \qquad \text{for} \qquad t = 1, 2, \ldots, j \neq 0
\end{aligned}
$$

Notice a White noise process is also covariance stationary since it satisfies all three conditions: the mean, variance and autocovariances are all finite and do not depend on time.

Conditional expectations and information set

The final important concepts are **conditional expectation** and **information set**.

The information set at time t is denoted $\mathcal{F}_t$ and contains all time t measurable events, and so the information set includes realization of all variables which have occurred on or before t.

For e.g., the information set for January 3rd 2021 contains all stock returns up to and including those which occurred on January 3rd.

Many expectations will often be made conditional on the time-t information set, expressed $E\left[y_{t+h}|\mathcal{F}_t\right]$, or in abbreviated form as $E_t\left[y_{t+h}\right]$.

Conditional variance is similarly defined, $V\left[y_{t+h}|\mathcal{F}_t\right] = V_t\left[y_{t+h}\right]$.

ARMA models

ARMA models

Autoregressive moving average (ARMA) processes form the core of time-series analysis.

The ARMA class can be decomposed into two smaller classes:

- Autoregressive (AR) processes.
- Moving average (MA) processes.

Moving average processes

The 1^{st} order moving average, written MA(1), is the simplest time-series process.

This process stipulates that the current value of y_t depends on both a new and the previous shock.

Definition (MA(1))

A first-order Moving Average process MA(1) has dynamics which follow:

$$y_t = \phi_0 + \theta_1 \epsilon_{t-1} + \epsilon_t,$$

where ϕ_0 and θ_1 are parameters and ϵ_t a white noise series with the additional property that $E_{t-1}[\epsilon_t] = 0$.

Moving average processes

The conditional and unconditional moments are defined as follows:

- The conditional mean is: $E_{t-1}\left[y_t\right] = \phi_0 + \theta_1 \epsilon_{t-1}$
- The unconditional mean is: $E\left[y_t\right] = \phi_0$
- The unconditional variance is: $V\left[y_t\right] = \sigma^2\left(1 + \theta_1^2\right)$
- The conditional variance is: $V_{t-1}\left[y_t\right] = \sigma_t^2$

The difference between the unconditional and the conditional moments reflects the persistence of the previous shock in the current period.

For instance, the unconditional variance is unambiguously larger than the average conditional variance which reflects the extra variability introduced by the moving average term.

Moving average processes

Finally, the autocovariance can be derived as,

- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-1} - E\left[y_{t-1}\right]\right)\right] = \theta_1 \sigma^2$
- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-2} - E\left[y_{t-2}\right]\right)\right] = 0$

By inspection of the autocovariance between y_t and y_{t-2} is follows that

$$\gamma_s = E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-2} - E\left[y_{t-2}\right]\right)\right] = 0 \qquad \text{for} \qquad s \geq 2,$$

Moving average processes

Definition (Moving average of order Q)

A moving average process of order Q, abbreviated MA(Q), has dynamics which follow

$$y_t \phi_0 + \sum_{q=1}^{Q} \theta_q \epsilon_{t-q} + \epsilon_t,$$

where ϵ_t is white noise series with the additional property that $E_{t-1}\left[\epsilon_t\right] = 0$.

The following properties hold in higher order moving averages:

- $E\left[y_t\right] = \phi_0$
- $V\left[y_t\right] = \left(1 + \sum_{q=1}^{Q} \theta_q^2\right) \sigma^2$
- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-2} - E\left[y_{t-2}\right]\right)\right] = \sigma^2 \sum_{i=0}^{Q-s} \theta_i \theta_{i+s}$ for $s \leq Q$
- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-2} - E\left[y_{t-2}\right]\right)\right] = 0$ for $s > Q$

Autoregressive processes

The other fundamental subclass of ARMA processes is the autoregressice process (AR)

Definition (First Order Autoregressive Process)

A first-order autoregressive process, abbreviated AR(1), has dynamics which follow

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t,$$

where ϵ_t is a white noise process with the additional property that $E_{t-1}\left[\epsilon_t\right] = 0$.

Autoregressive processes

Unlike the MA(1) process, y appears on both sides of the equation. However, this is only convenience and the process can be recursively substituted to provide an expression that depends only on the errors ϵ_t and an initial condition

$$y_t \sum_{i=0}^{t-1} \phi_1^i \phi_0 + \sum_{i=0}^{t-1} \phi_1^i \epsilon_{t-i} + \phi_1^t y_0,$$

In many cases the initial condition is irrelevant, especially for large samples.

As long as $|\phi_1| < 1$, $\lim_{t \to \infty} \phi^t y_0 \to 0$ and the effect of an initial condition will be small. Assuming $|\phi_1| < 1$ the solution simplifies to

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \epsilon_t, \\ y_t &= \sum_{i=0}^{\infty} \phi_1^i \phi_0 + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-1}, \\ y_t &= \frac{\phi_0}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \end{aligned}$$

where the identity $\sum_{i=0}^{\infty} \phi_1^i = (1-\phi_1)^{-1}$ is used in the final solution.

Autoregressive processes

The conditional and unconditional moments are defined as follows:

- The conditional mean is: $E_{t-1}\left[y_t\right] = \phi_0 + \phi_1 y_{t-1}$
- The unconditional mean is: $E\left[y_t\right] = \frac{\phi_0}{1-\phi_0}$
- The unconditional variance is: $V\left[y_t\right] = \frac{\sigma^2}{1-\phi_1^2}$
- The conditional variance is: $V_{t-1}\left[y_t\right] = E_{t-1}\left[\epsilon_t^2\right] = \sigma_t^2$

The difference between the unconditional and the conditional moments reflects the persistence of the previous shock in the current period.

For instance, the unconditional variance is unambiguously larger than the average conditional variance and the variance explodes as $|\phi_1|$ approaches 1 or -1.

Finally, the autocovariance can be derived as,

- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-2} - E\left[y_{t-2}\right]\right)\right] = \phi_1^s \frac{\sigma^2}{1-\phi_1^2}$

The AR(1) can be extended to the AR(P) class by including additional lags of y_t.

Definition (Autoregressive process of order P)

An Autoregressive process of order P AR(P) has dynamics which follow

$$y_t = \phi_0 + \sum_{p=1}^{P} \phi_p y_{t-p} + \epsilon_t,$$

where ϵ_t is white noise series with additional property that $E_{t-1}\left[\epsilon_t\right] = 0$.

Autoregressive processes

Some of the more useful properties of general AR process are:

- $E\left[y_t\right] = \frac{\phi_0}{1-\sum_{p=1}^{P} \phi_p}$
- $V\left[y_t\right] = \frac{\sigma^2}{1-\sum_{p=1}^{P} \phi_p \rho_p}$ where ρ_p is the p^{th} autocorrelation.
- $V\left[y_t\right]$ is infinite if $\sum_{p=1}^{P} \phi_p \geq 1$
- $E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-s} - E\left[y_{t-s}\right]\right)\right] \neq 0$ for any s

This four properties point to some important regularities of AR processes.

1. The mean is only finite if $\sum_{p=1}^{P} \phi_p < 1$.
2. Autocovariances are (generally) not zero, unlike for MA processes.

Autoregressive-moving average processes

Putting these two processes together yields the complete class of ARMA processes.

Definition (Autoregressive-moving average process)

An autoregressive moving average process with orders P and Q (ARMA(P,Q)) has dynamics which follow

$$y_t = \phi_0 + \sum_{p=1}^{P} \phi_p y_{t-p} + \sum_{q=1}^{Q} \theta_q \epsilon_{t-q} + \epsilon_t,$$

where ϵ_t is a white noise process with the additional property that $E_{t-1}[\epsilon_t] = 0$.

Autoregressive-moving average processes

Let consider the simplest ARMA(1,1) process that includes a constant term

$$y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t,$$

To derive the properties of this model it is useful to convert the ARMA(1,1) into its infinite lag representation using recursive substitution, i.e.,

$$y_t = \frac{\phi_0}{1-\phi_1} + \epsilon_t + \sum_{i=0}^{\infty} \phi_1^i \left(\phi_1 + \theta_1\right) \epsilon_{t-i-1},$$

Using the infinite-lag representation, the unconditional and conditional means can be computed,

- $E\left[y_t\right] = \frac{\phi_0}{1-\phi_1}$
- $E_{t-1}\left[y_t\right] = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1}$

Since y_{t-1} and ϵ_{t-1} are in the time $t-1$ information set, these variables pass through the conditional expectation.

Stationarity of ARMA models

Stationarity conditions for ARMA processes can be determined using lag polynomials, i.e.,

$$
\begin{aligned}
y_t &= \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t, \\
y_t - \phi_0 - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p} &= \phi_0 + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t, \\
\left(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p\right) y_t \phi_0 + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t,
\end{aligned}
$$

This is a linear difference equation, and the stability conditions depend on the roots of the characteristic polynomial

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \ldots - \phi_{p-1} z - \phi_p$$

In the simple AR(1) case, this corresponds to $|z_1| < 1$.

For higher order models, stability must be checked by numerically solving the characteristic equation.

Example: Stock returns and default risk

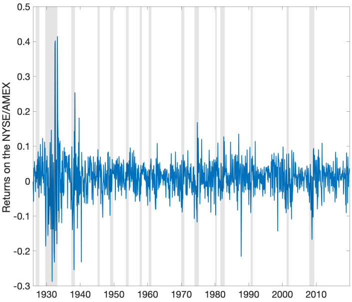
Two series will be used to investigate the properties of stationary time series:

- Returns on the value-weighted stock market index
- Spread between the average interest rates on portfolios of Aaa-rated and Baa-rated corporate bonds, commonly known as the default spread or default premium.

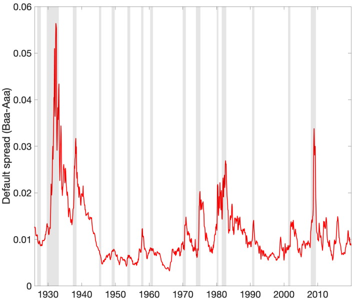
Both the returns and the credit spread are available from January 1927 to January 2021.

Example: Stock returns and default risk

The dynamics of the equity and default risk premiums from 1925 to 2019.



(a) Market returns



(b) Default premium

Example: Stock returns and default risk

The dynamics of the equity and default risk premiums from 1925 to 2019.

Model	VWM				Baa-Aaa			
	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\sigma}$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\sigma}$
AR(1)	0.009	0.085		0.003	0.001	0.975		0.001
	(0.002)	(0.016)		(0.001)	(0.000)	(0.003)		(0.000)
MA(1)	0.009		0.086	0.003	0.011		0.894	0.001
	(0.002)		(0.015)	(0.001)	(0.001)		(0.005)	(0.000)
ARMA(1,1)	0.009	-0.007	0.093	0.003	0.003	0.964	0.241	0.001
	(0.003)	(0.198)	(0.199)	(0.001)	(0.001)	(0.004)	(0.012)	(0.000)

Parameters estimates and standard errors in parenthesis.

Autocorrelation and partial autocorrelations

Autocorrelations and the autocorrelation function

Autocorrelations are to autocovariances as correlations are to covariances. That is, the s^{th} autocorrelation is the s^{th} autocovariance divided by the product of the variance of y_t and y_{t-s}, and when a process is covariance stationary, $V[y_t] = V[y_{t-s}]$, and so $\sqrt{V[y_t]\,V[y_{t-s}]} = V[y_t]$.

Definition (Autocorrelation)

The autocorrelation of a covariance stationary scalar process is defined

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t-s} - E\left[y_{t-s}\right]\right)\right]}{V\left[y_t\right]},$$

where γ_s is the s^{th} autocovariance.

Autocorrelations and the autocorrelation function

The autocorrelation function (ACF) relates the lag length (s) and the parameters of the model to the autocorrelation.

Definition (Autocorrelation function)

The autocorrelation function (ACF), $\rho(s)$, is a function of the population parameters that defines the relationship between the autocorrelations of a process and the lag length.

The variance of a covariance stationary AR(1) is $\sigma^2\left(1-\phi^2\right)^{-1}$ and the s^{th} autocovariance is $\phi^s\sigma^2\left(1-\phi^2\right)^{-1}$, and so the ACF is

$$\rho(s)=\frac{\phi^s\sigma^2\left(1-\phi^2\right)^{-1}}{\sigma^2\left(1-\phi^2\right)^{-1}}=\phi^s,$$

Partial autocorrelation and the partial autocorrelation function

Partial autocorrelations are simiular to autocorrelations with one important difference: the s^{th} partial autocorrelation still relates y_t and y_{t-s} but it eliminates the effects of $y_{t-1}, y_{t-2}, \ldots, y_{t-(s-1)}$.

Definition (Partial autocorrelation)

The s^{th} partial autocorrelation (φ_s) is defined as the population value of the regression coefficient on ϕ_s in

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{s-1} y_{t-(s-1)} + \phi_s y_{t-s} + \epsilon_t,$$

Like the autocorrelation function, the partial autocorrelation function (PACF) relates the partial autocorrelation to population parameters and lag length.

Partial autocorrelation and the partial autocorrelation function

The partial autocorrelations are interpretable as regression coefficients.

For example, in a stationary AR(1) model, $y_t = \phi_1 y_{t-1} + \epsilon_t$, the PACF is

$$\begin{aligned} \varphi(s) &= \phi_1^{|s|}, \qquad s = 0, 1, -1, \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Notice:

- That $\varphi_0 = \phi^0 = 1$ is obvious: the correlation of a variable with itself is 1.
- The first partial autocorrelation is defined as ϕ_1 in the regression $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$, i.e., $\varphi_1 = \phi_1$
- The second partial autocorrelation is defined as ϕ_2 in the regression $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, i.e., $\varphi_2 = 0$ since y_{t-2} has no effect on y_t once y_{t-1} is included.

Partial autocorrelation and the partial autocorrelation function

The behavior of the ACF and PACF for various members of the ARMA family can be summarised as follows:

Process	ACF	PACF
White noise	All 0	All 0
AR(1)	$\rho_s = \phi^s$	0 beyond lag 2
AR(P)	Decays toward zero exponentially	Non-zero through lag P, 0 thereafter
MA(1)	$\rho_1 \neq 0, \rho_s = 0$ for $s > 1$	Decays toward zero exponentially
MA(Q)	Non-zero through lag Q, 0 thereafter	Decays toward zero exponentially
ARMA(P,Q)	Exponential decays	Exponential decay

Sample autocorrelations and partial autocorrelations

Sample autocorrelations are computed using sample analogues of the population moments in the definition of an autocorrelation.

Define $y_t^* = y_t - \overline{y}$ to be the demeaned series where $\overline{y} = T^{-1} \sum_{t=1}^{T} y_t$. The s^{th} sample autocorrelation is defined

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{T} y_t^* y_{t-s}^*}{\sum_{t=1}^{T} (y_t^*)},$$

A plot of the sample autocorrelations against the lag index is known as a sample

autocorrelogram

Under the null $H_0 : \rho_s = 0, s \neq 0$, inference can be made under homoskedasticity noting that $V[\hat{\rho}_s] = T^{-1}$ using a standard t-test

$$T^{-1/2} \hat{\rho}_s \xrightarrow{d} N(0,1),$$

Sample autocorrelations and partial autocorrelations

Partial autocorrelations can be estimated using regressions,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \hat{\varphi}_s y_{t-s} + \epsilon_t,$$

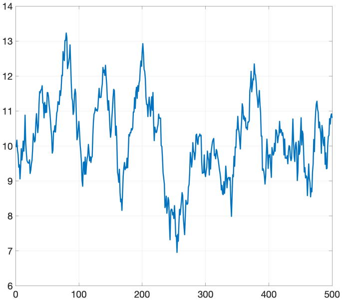
where $\varphi_s = \hat{\phi}_s$.

To test whether a partial autocorrelation is zero, the variance of $\hat{\phi}_s$, under the null and assuming homoskedasticity, is approximately T^{-1} for any s, and so a standard t-test can be used,

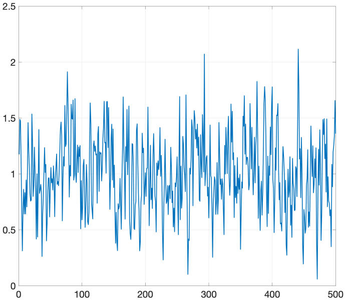
$$T^{-1/2}\hat{\phi}_s \xrightarrow{d} N(0,1),$$

A plot of the sample partial autocorrelations against the lag index is know as a sample partial autocorrelogram.

Sample autocorrelations and partial autocorrelations



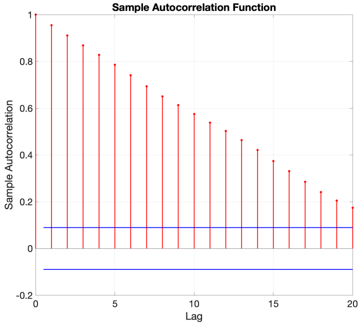
(c) AR(1) $\phi_1 = 0.9$



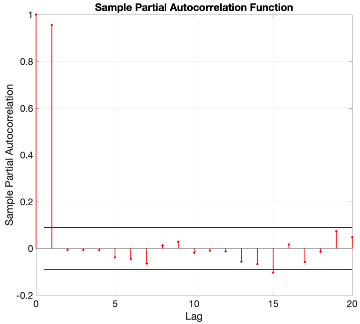
(d) AR(1) $\phi_1 = 0.5$

Sample autocorrelations and partial autocorrelations

AR(1) with $\phi_1 = 0.9$



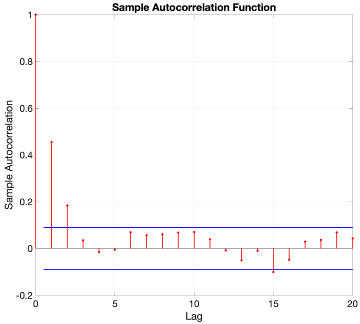
(e) ACF



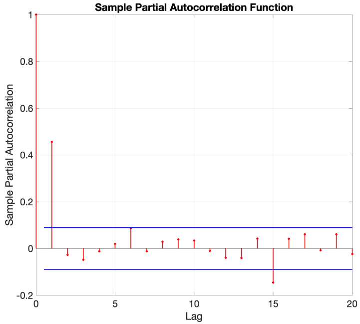
(f) PACF

Sample autocorrelations and partial autocorrelations

AR(1) with $\phi_1 = 0.5$



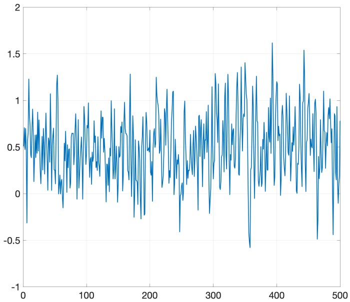
(g) ACF



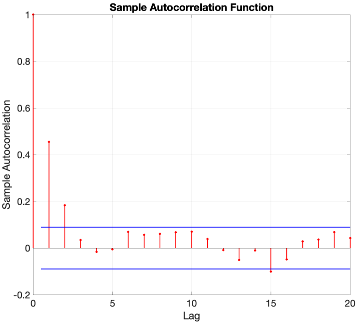
(h) PACF

Sample autocorrelations and partial autocorrelations

MA(1) with $\theta_1 = 0.5$



(i) MA(1) with $\theta_1 = 0.5$



(j) ACF

Sample autocorrelations and partial autocorrelations

Tests that multiple autocorrelations are simultaneously zero can also be conducted. The standard method to test that s autocorrelations are zero, $H_0 : \rho_1 = \rho_2 = \ldots = \rho_s = 0$, is the Ljung-Box Q statistic.

Definition (Ljung-Box Q statistic)

The Ljung-Box Q statistic, or simply Q statistic, tests the null that the first s autocorrelations are all zero against an alternative that at least one is non-zero. The test statistic is defined as:

$$Q = T(T+2)\sum_{k=1}^{s}\frac{\hat{\rho}_k^2}{T-k},$$

and Q has a standard χ^2_s distribution.

The Q statistic is only valid under an assumption of homoskedasticity so caution is warranted when using it with financial data.

A heteroskedasticity robust version of the Q-stat can be formed using an LM test.

Definition (LM test for serial correlation)

Under the null, $E\left[y_t^* y_{t-j}^*\right] = 0$ for $1 \leq j \leq s$. The LM-test for serial correlation is constructed by defining the score vector $\mathbf{s}_t = y_t^* \left[y_{t-1}^* y_{t-2}^* \ldots y_{t-s}^*\right]'$,

$$LM = T\bar{\mathbf{s}}'\hat{\mathbf{S}}\bar{\mathbf{s}} \xrightarrow{d} \chi_s^2,$$

where $\bar{\mathbf{s}} = T^{-1}\sum_{t=1}^{T}\mathbf{s}_t$ and $\hat{\mathbf{S}} = T^{-1}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_t'$.

Estimation

Choosing the model specification: The Box-Jenkins approach

The Box and Jenkins methodology is the most common approach for time-series model selection. It consists of two stages:

- **Identification:** Visual inspection of the series, the autocorrelations and the partial autocorrelations.
- **Estimation:** By relating the ACF and PACF of ARMA models, candidates models are identified.

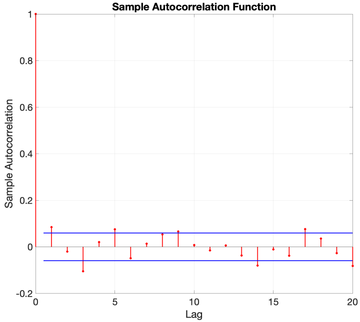
Models are typically selected through information criteria:

- **Akaike Information Criterion:** The AIC is defined as $AIC = \ln \hat{\sigma}^2 + \frac{2k}{T}$ where $\hat{\sigma}^2$ is the estimated variance of the regression error and k the number of parameters.
- **Bayesian Information Criterion:** The BIC is defined as $BIC = \ln \hat{\sigma}^2 + \frac{k \ln T}{T}$ where $\hat{\sigma}^2$ is the estimated variance of the regression error and k the number of parameters.

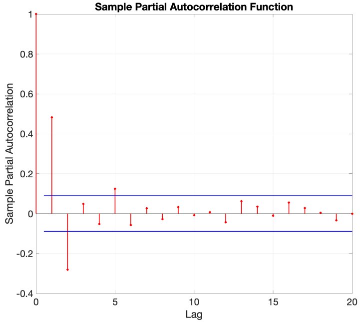
Notice ICs are often applied by estimating the largest model then dropping lags until the AIC/BIC fail to decrease.

Example: Stock returns and default risk

Aggregate stock market returns



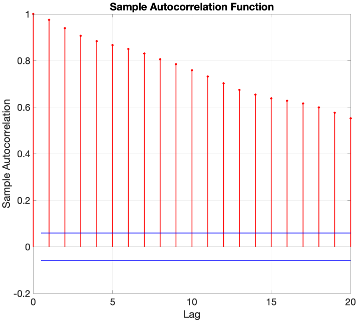
(k) ACF



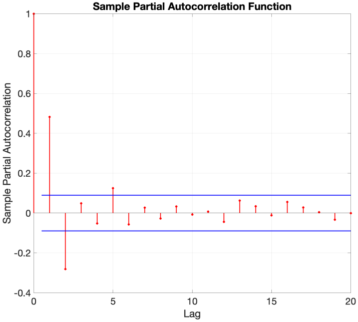
(l) PACF

Example: Stock returns and default risk

Default premium



(m) ACF



(n) PACF

Estimation

ARMA models are typically estimated using maximum likelihood (ML) estimation assuming that errors are normally distributed.

One way to estimate the model parameters is the so-called **conditional maximum likelihood**, where the likelihood of y_t given $y_{t-1}, y_{t-2}, \ldots$ is used.

The data are assumed to be conditionally normal, and so the likelihood is:

$$f\left(y_t | y_{t-1}, y_{t-2}, \ldots; \phi, \theta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right),$$

$$= \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{\left(y_t - \phi_0 - \sum_{i=1}^{P} \phi_i y_{t-i} - \sum_{j=1}^{Q} \theta_j \epsilon_{t-j}\right)^2}{2\sigma^2}\right),$$

Estimation

Since the $\{\epsilon_t\}$ series is assumed to be a white noise process, the joint likelihood is simply the product of the individual likelihoods,

$$f\left(\mathbf{y}_t|\mathbf{y}_{t-1},\mathbf{y}_{t-2},\ldots;\phi,\theta,\sigma^2\right)=\prod_{t=1}^{T}\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right),$$

and the conditional log-likelihood is given by

$$l\left(\phi,\theta,\sigma^2;\mathbf{y}_t|\mathbf{y}_{t-1},\mathbf{y}_{t-2},\ldots\right)=-\frac{1}{2}\sum_{t=1}^{T}\ln 2\pi+\ln\sigma^2+\frac{\epsilon_t^2}{\sigma^2},$$

The conditional maximum likelihood estimation of ARMA models requires either backcast values or truncation since some of the observations have low indices that depend on observations not in the sample.

The conditional likelihood works well when data are not overly persistent and T is not too small.

Estimation

Recall that the first-order condition for the mean parameters from a normal log-likelihood does not depend on σ^2 and that given the parameters in the mean equation, the maximum likelihood estimate of the variance is

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \epsilon_t^2,$$

This allows the variance to be concentrated out of the log-likelihood so that it becomes

$$l\left(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots; \phi, \theta, \sigma^2\right) = -\frac{1}{2} \sum_{t=1}^{T} \ln 2\pi - \frac{T}{2} - \frac{T}{2} \ln \hat{\sigma}^2,$$

Then, eliminating terms that do not depend on model parameters shows that maximizing the likelihood is equivalent to minimizing the error variance, i.e.,

$$\max_{\phi,\theta,\sigma^2} l\left(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots; \phi, \theta, \sigma^2\right) = -\frac{T}{2} \ln \hat{\sigma}^2,$$

Thus, estimation via conditional MLE is equivalent to least squares.

Exercises

Exercises

Problem 5.1: Under what conditions on the parameters and errors are the following processes covariance stationary?

- $y_t = \phi_0 + \epsilon_t$
- $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$
- $y_t = \phi_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$
- $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$

Problem 5.2: consider an AR(1) $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$. What are the values of

- $E\left[y_{t+1}\right]$
- $E_t\left[y_{t+1}\right]$
- $V\left[y_{t+1}\right]$
- $V_t\left[y_{t+1}\right]$

Exercises

Problem 5.3: Consider an MA(1) $y_t = \phi_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$. What are the values of

- $E[y_{t+1}]$
- $E_t[y_{t+1}]$
- $V[y_{t+1}]$
- $V_t[y_{t+1}]$

Problem 5.4: For each of the following processes, find $E_t[y_{t+1}]$. You can assume ϵ_t is an i.i.d. sequence with mean zero.

- $y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$
- $y_t = \phi_0 + \theta_1 y_{t-1} + \epsilon_t$
- $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$
- $y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$

Problem 5.5: How are the autocorrelations and the partial autocorrelations useful in building a model?

Problem 5.6: Suppose you observe the three sets of ACF/PACF in the simulation above. What ARMA specification would you expect in each case?

Problem 5.7: Describe the information criteria for model selection.

Problem 5.8: Suppose $y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$ where $\{\epsilon_t\}$ is a white noise process.

- Describe the two concepts of stationarity
- Why is stationarity a useful property?
- What conditions on the model paremeters are needed for $\{y_t\}$ to be covariance stationary?
- Describe the Box-Jenkins methodology for model selection