STAT 5170: Applied Time Series

Course notes for part B of learning unit 2

Section 2B.1: Stationarity.

In developing a conceptual understanding of time series models, it is helpful to define a sense by which a time series would be in a state of stable, steady evolution. In other words, it is in a state where it is not drifting to ever-increasing regions of values, and the variability it exhibits is manageable. The following two definitions offer strong and weak versions of a concept that suitably captures the idea of a time series evolving in a steady state:

Definition: A time series (x_t) is *strongly (a.k.a. strictly) stationary* if its joint distributions are invariant to time shifts. That is, the time series is strictly stationary if for any time sequence $t_1, \ldots, t_k$, time shift h, and constants $c_1, \ldots, c_k$, it is the case that

$$P[x_{t_1} \leq c_1, \ldots, x_{t_k} \leq c_k] \quad = \quad P[x_{t_1+h} \leq c_1, \ldots, x_{t_k+h} \leq c_k].$$

Definition: A time series (x_t) is *weakly stationary* if

(*i.*) $\mu_t = \mu$, a constant across t

(*ii.*) $\gamma(s,t)$ depends only on $|s-t|$.

As we can see, the strong version of stationarity refers to distributional properties of a time series, and is more completely captures the concept. The weak version refers only to means and covariances, and is sufficient for simpler modes of inference. As for relationships between them, it is straightforward to deduce that strong stationarity implies weak stationarity.

Weak stationarity does not necessarily imply strong stationarity. However, it would if the time series is Gaussian. To state such a result, it is convenient to first define the following:

Definition: A time series (x_t) is *Gaussian* if its joint distributions are normal. That is, the time series is Gaussian if for every $t_1, \ldots, t_k$, the vector $\boldsymbol{x} = [t_1, \ldots, t_k]^T$ has a multivariate normal distribution.

With this definition at hand, it is readily deduced that a Gaussian time series (x_t) is strongly stationarity if it is weakly stationarity. This follows from the elementary property of a multivariate normal distribution by which the entire distribution is determined from its mean vector and covariance matrix. In what follows, if neither version of stationarity is specified nor is a Gaussian assumption specified or implied from the context, then weak stationarity is to be assumed.

When the stationarity of a time series is established, it is convenient to adopt a simplified notation for autocovariance and autocorrelation functions:

Notation: If (x_t) is stationary, denote by $\gamma(h)$ the autocovariance value $\gamma(t+h,t) = Cov[x_{t+h}, x_t]$, and by $\rho(h)$ the autocorrelation value $\rho(t+h,t) = Corr[x_{t+h}, x_t]$.

Using this notation, and assuming stationarity, a few basic properties are readily deduced and stated without proof:

Properties: Suppose (x_t) is stationary. It follows that...

(*i.*) $\gamma(h)$ and $\rho(h)$ are well defined, since, for any t, $\gamma(t+h,t)$ depends only on h;

(*ii.*) $\gamma(s,t) = \gamma(|s-t|)$;

(*iii.*) $\gamma(h) = \gamma(t+h,t) = \gamma(t,t-h) = \gamma(t-h,t) = \gamma(-h)$;

(*iv.*) and, because $\gamma(h) = \gamma(h,0)$ and $\gamma(0) = \gamma(0,0) = \gamma(h,h)$, autocorrelation may be understood through the formula

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

□

Example: Stationarity of white noise

Suppose the time series (w_t) is Gaussian white noise. Its mean function and autocovariance function is

$$\mu_t = 0 \quad \text{and} \quad \gamma(s,t) = \begin{cases} \sigma_w^2 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases}$$

It is seen that the mean function is constant and the autocovariance function depends only on whether $s - t = 0$ or $s - t \neq 0$. The time series is therefore weakly stationary. □

Example: Non-stationarity of a random walk

Suppose the time series (x_t) is a Gaussian random walk, with initial state x_0, and drift parameter δ. Its mean function and autocovariance function is

$$\mu_t = x_0 + \delta t \quad \text{and} \quad \gamma(s,t) = \min\{s,t\}\sigma_w^2$$

The mean function is constant only if $\delta = 0$, so the only hope of stationarity is when there is no drift. However, even in that case, the autocovariance function does not satisfy the condition of stationarity. In particular, we know that the variance of a random walk increases with time,

$$Var[x_t] = \gamma(t,t) = t\sigma_w^2.$$

which means that, if $s \neq t$,

$$\gamma(s,s) = s\sigma_w^2 \neq t\sigma_w^2 = \gamma(t,t)$$

even though in each of $\gamma(s,s)$ and $\gamma(t,t)$ the difference between two arguments of the autocovariance function is the same (specifically, the difference is zero). □

Example: Possible stationarity of $AR(1)$ autoregressive time series

We have learned that a random walk with initial value x_0, defined by the relationship

$$x_t = \delta + x_{t-1} + w_t, \text{ for } t = 1, 2, \ldots,$$

has the representation

$$x_t = \delta t + x_0 + \sum_{i=1}^{t} w_i.$$

Subsequently, if x_0 is fixed, the mean and variance of x_t are quickly deduced as $E[x_t] = \delta t + x_0$ and $Var[x_t] = \sigma_w^2 t$.

A similar line of deduction may be used to work with a $AR(1)$ autoregressive time series with a fixed initial value, x_0. Suppose the time series (x_t) satisfies the relationship

$$x_t = \phi_1 x_{t-1} + w_t, \text{ for } t = 1, 2, \ldots,$$

where ϕ_1 is a parameter, and (w_t) is Gaussian white noise. An associated representation is

$$x_t = \phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i}, \text{ for } t = 1, 2, \ldots,$$

A proof is by induction: (*i.*) At $t = 1$, the definition of the $AR(1)$ relationship provides $x_1 = \phi_1 x_0 + w_1$ (*ii.*) For $t > 1$, apply the induction hypothesis to see that

$$\begin{aligned} x_t &= \phi_1 \left\{ \phi_1^{t-1} x_0 + \sum_{i=0}^{t-2} \phi_1^i w_{t-1-i} \right\} + w_t \\ &= \phi_1^t x_0 + w_t + \sum_{i=0}^{t-2} \phi_1^{i+1} w_{t-1-i} \\ &= \phi_1^t x_0 + w_t + \sum_{i=1}^{t-1} \phi_1^i w_{t-i} \\ &= \phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i}. \end{aligned}$$

Using this expression, and the properties of white noise, when x_0 is fixed, the mean and variance of an individual measurement are deduced as

$$E[x_t] = E\left[\phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i} \right] = \phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i E[w_{t-i}] = \phi_1^t x_0$$

and

$$Var[x_t] = Var\left[\phi_1^t x_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i} \right] = \sum_{i=0}^{t-1} \phi_1^{2i} Var[w_{t-i}] = \sum_{i=0}^{t-1} \phi_1^{2i} \sigma_w^2.$$

The variance formula may be refined further by applying the geometric series formula

$$\sum_{i=0}^{n-1} u^i \;=\; \frac{1-u^n}{1-u}, \text{ for } u \neq 1$$

The variance is then

$$Var[x_t] \;=\; \frac{1-\phi_1^{2t}}{1-\phi_1^2}\sigma_w^2, \text{ for } |\phi_1| \neq 1$$

The case $\phi_1 = 1$ is equivalent to a random walk.

Because the variance depends on t, the time series is not stationary. However, part of the reason for this that the initial value, x_0, is fixed. To treat x_0 as random, one perspective is to observe the time series properties when t is large, which is roughly equivalent to sending x_0 into the infinite past, where its influence on the time series is negligible.

This perspective indicates that, when $|\phi_1| < 1$, the variance converges to

$$Var[x_t] \;\approx\; \frac{1}{1-\phi_1^2}\sigma_w^2.$$

On the other hand, when $|\phi_1| > 1$ the variance formula diverges to infinity (when $\phi_1 > 1$), or, at any rate, fails to converge (*e.g.*, it oscillates across an increasing range when $\phi_1 < -1$).

The indication, here, is that the only hope for stationarity of an $AR(1)$ autoregressive time series is under the setting $|\phi_1| < 1$. □

Section 2B.2: Autocorrelation statistics.

When analyzing data from a time series that is thought to be stationary, a few simple, fairly intuitive summary-statistic formulas are available to provide sample versions of the series's constant mean value, μ, as well as its autocovariance and autocorrelation functions, $\gamma(h)$ and $\rho(h)$:

Definition: Suppose $x_1, \ldots, x_n$ is a stationary time series observed on a finite time span. Then, for $h = 0, \ldots, n-1$,

(*i.*) the *sample mean* is

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t;$$

(*ii.*) the *sample autocovariance function* is

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

(*iii.*) and, the *sample autocorrelation function* (SACF) is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

An elementary property of the sample mean is that its expected value is μ, the time series's constant mean value; that is $E[\bar{x}] = n^{-1}\sum_{t=1}^{n} E[x_t] = n^{-1}\sum_{t=1}^{n}\mu = \mu$. Asymptotic theory of stationary time series provides a number of large-sample distributional properties of the statistics, such as the following:

Properties:

(*i.*) As $n \to \infty$, the sample mean converges in probability to time series's constant mean value; that is $\bar{x} \to_p \mu$.

(*ii.*) Stationarity implies that the distribution of

$$y_t = (x_{t+h} - \bar{x})(x_t - \bar{x})$$

does not depend on t. Large-sample theory subsequently provides that $\hat{\gamma}(h) \to_p \gamma(h)$ as $n \to \infty$.

(*iii.*) Similarly, $\hat{\rho}(h) \to_p \rho(h)$ as $n \to \infty$.

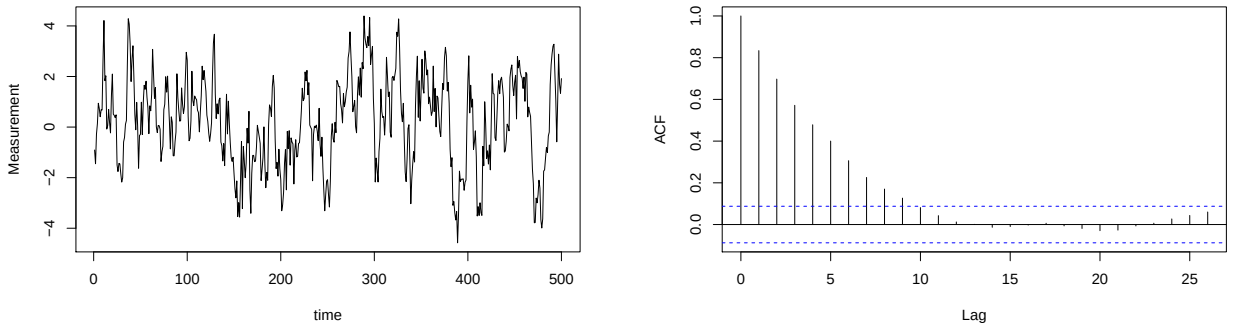
(*iv.*) Large-sample theory also provides that when (x_t) is white noise and n is large, the sample ACF, $\hat{\rho}(h)$, is, for each $h > 0$, approximately normal with mean zero and standard deviation $1/\sqrt{n}$.

These properties are sometime cited as the basis of a certain approach to statistical inference methodology whose motivating concept is *hypothetical repeated sampling* and the associated concept of the *sampling distribution of a statistic*. This approach of statistical inference is likely what was taught in your earlier classes. Within such methodology, $\bar{x}$, $\hat{\gamma}(h)$, and $\hat{\rho}(h)$ are said to be *asymptotically consistent estimates* of μ, $\gamma(h)$, and $\rho(h)$, respectively. The large-sample normal approximation to the distribution of the sample ACF is taken to suggest a rejection rule for a *hypothesis test* of $H_0 : \rho(h) = 0$ *vs* $H_0 : \rho(h) \neq 0$; the rule is to reject H_0 precisely when $|\hat{\rho}(h)| > 2/\sqrt{n}$, which implies a *type I error probability* of about $\alpha = 0.05$.

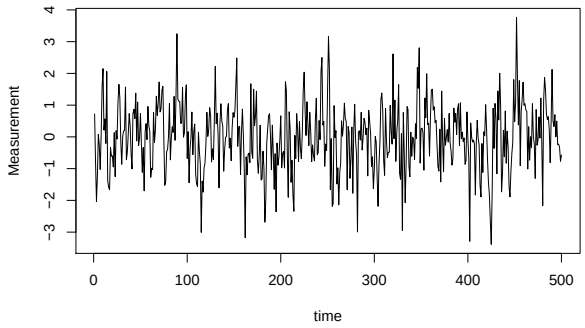
Example: Sample ACF of simulated $AR(1)$ time series

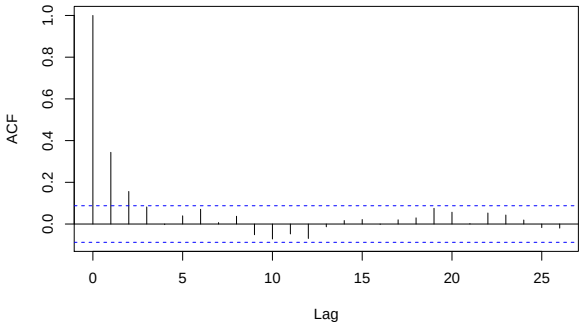
The following figures display sample paths and sample ACFs from $AR(1)$ models with ϕ_1 set to a range of values. Each sample path is composed of $n = 500$ simulated measurements.

A sample path and sample ACF for the case $\phi_1 = 0.9$ is as follows.

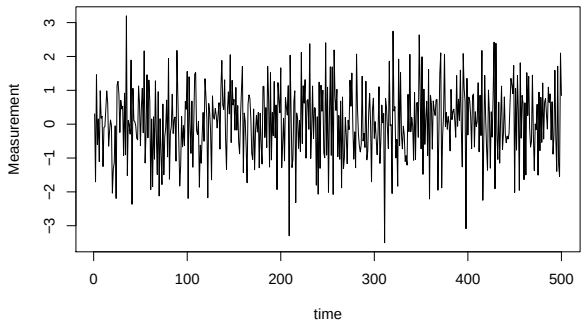


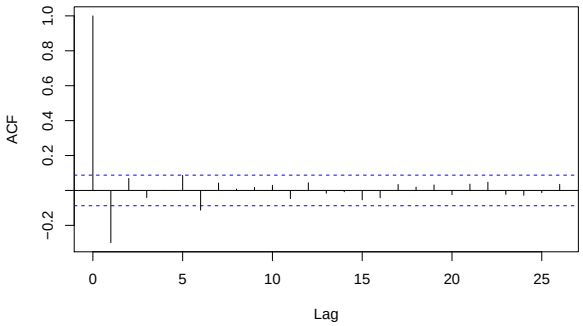
For the case $\phi_1 = 0.3$, these are



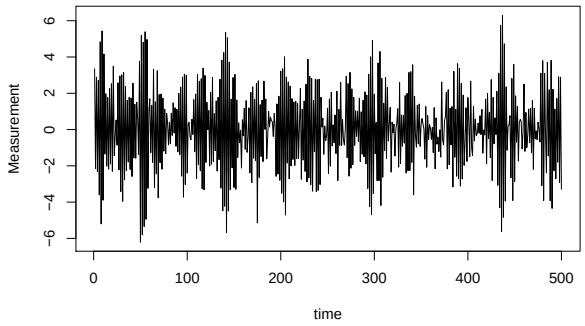


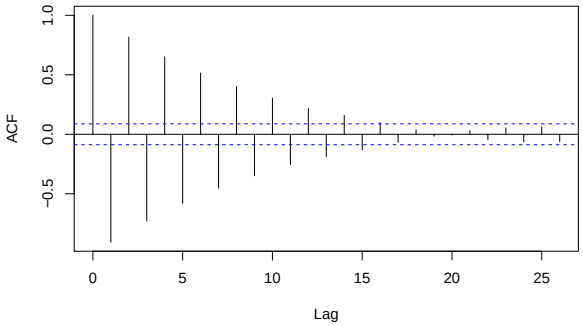
For $\phi_1 = -0.3$, they are





For $\phi_1 = -0.9$, they are





□

Example: Sample ACF of a simulated $MA(2)$ time series

Suppose (x_t) is an $MA(q)$ moving-average time series with $q = 2$, $\theta_1 = -0.7$, and $\theta_2 = 0.5$, which is defined by the relationship

$$x_t = w_t - 0.7w_{t-1} + 0.5w_{t-2},$$

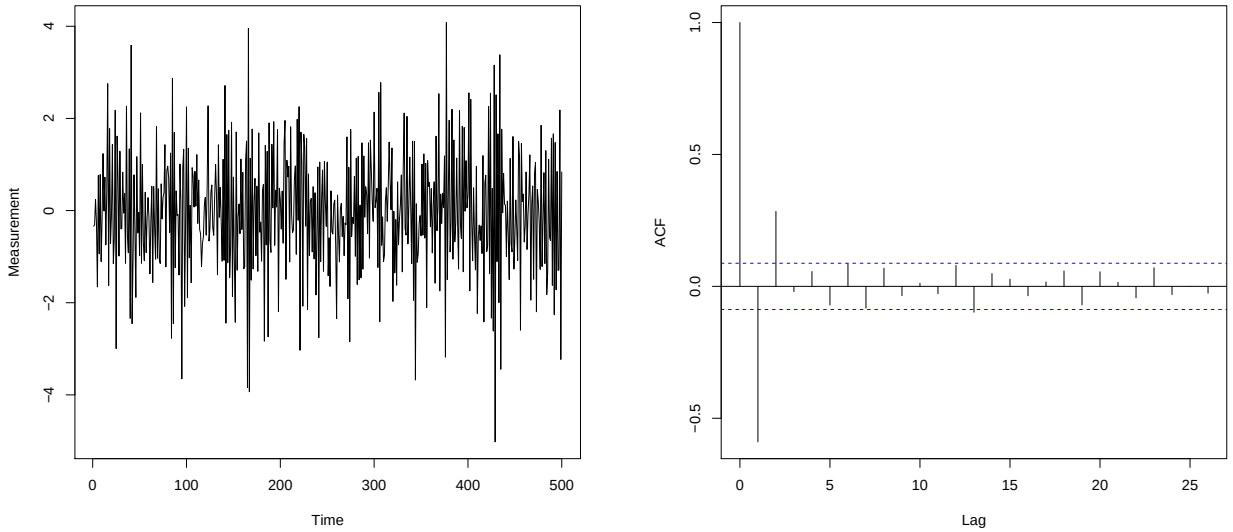
where (w_t) is Gaussian white noise with $\sigma_w^2 = 1$. From previous work with $MA(2)$ time series, we know that the mean function of this time series is $\mu_t = 0$, and autocovariance function of this time series depends on $|s - t|$ only; the time series is therefore stationary. Using special notation for stationary time series, its autocovariance function may be written

$$\gamma(h) = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma_w^2 = 1.74 & \text{if } h = 0 \\ \theta_1(1 + \theta_2)\sigma_w^2 = -1.05 & \text{if } h = 1 \\ \theta_2\sigma_w^2 = 0.5 & \text{if } h = 2 \\ 0 & \text{otherwise} \end{cases}$$

The ACF of this time series is

$$\rho(h) \quad = \quad \gamma(h)/\gamma(0) = \begin{cases} 1.0000 & \text{if } s = t \\ -0.6034 & \text{if } |s-t| = 1 \\ 0.2874 & \text{if } |s-t| = 2 \end{cases}$$

A time series plot and sample ACF of $n = 500$ simulated measurements from this model are as follows.



The sample ACF in the right panel exhibits the characteristic pattern for $MA(q)$ time series, which is that autocorrelations suddenly vanish when the lag value exceeds the parameter q. Because we are working with sample values, the autocorrelations do not vanish, but fall below the statistical cutoff values of

$$\text{cutoff} \quad = \quad 2/\sqrt{n} = 2/\sqrt{500} = 0.0894.$$

□

Besides their role in inference, the SACF can also play an important diagnostic role in time series analysis, by which it would help to identify a class of time series models, such as the $AR(p)$ or $MA(q)$ models, that may be suitable as a basis of inferential analysis of a set of time series data. The idea is to match patterns observed in the sample statistic to theoretical patterns that are deduced for a model's underlying ACF. We have not yet deduced what these theoretical patterns are, but a hint of the pattern associated with the $MA(q)$ models is given in the last example. A summary table of a few other patterns is as follows.

Model	ACF
white noise	0
trend	slow decay
periodic	periodic
$MA(q)$	0 for $h > q$
$AR(p)$	exponential decay to 0

Some of these are illustrated in the next example.

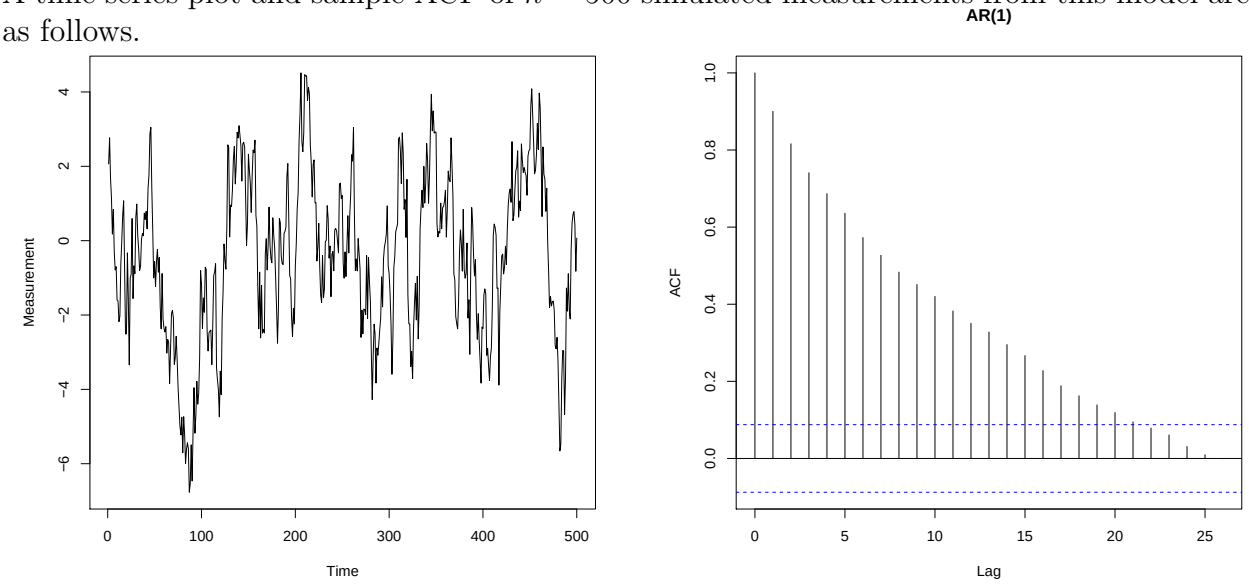
Example: Sample ACFs of other time series

Suppose (x_t) is an $AR(p)$ autoregressive time series with $p = 1$ and $\phi_1 = 0.9$, which is defined by the relationship

$$x_t = 0.9x_{t-1} + w_t,$$

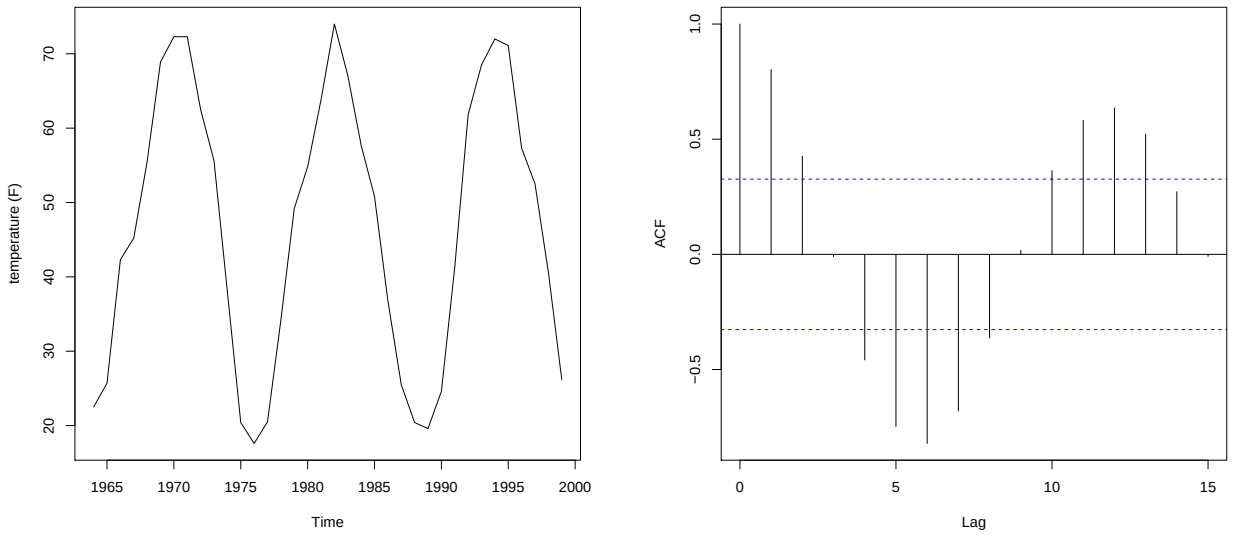
where (w_t) is Gaussian white noise with $\sigma_w^2 = 1$

A time series plot and sample ACF of $n = 500$ simulated measurements from this model are as follows.

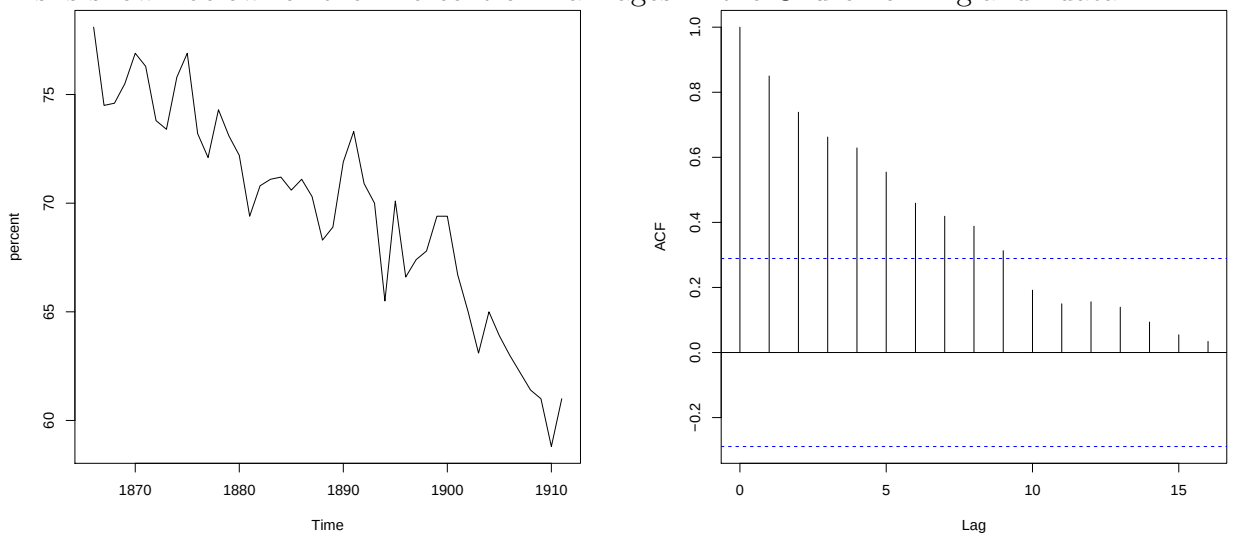


Observe in the left panel, depicting the sample path, that the time series appears "smooth" in the sense of outlining large, sweeping shapes. The sample ACF in the right panel exhibits the characteristic pattern for autoregressive time series, which is that the autocorrelations decay to zero at an exponential rate. We will derive the ACF of autoregressive time series later.

When the time series is strongly seasonal or periodic, the sample ACF will also exhibit a periodic pattern. This is shown below for the monthly temperature data for Dubuque, IA.



When a time series is nonstationary, the sample ACF is not sensible to calculate. Nevertheless, it can be calculated, and if it is, a plot of it will display slowly decaying autocorrelations. This is shown below for the "Percent of Marriages in the Church of England" data.



This plot shows a decay pattern similar to that in the sample ACF of the $AR(p)$ autoregressive time series, but here the decay is slower. □

Section 2B.3: Linear prediction and partial autocorrelation.

A concept in addition to autocorrelation, called *partial autocorrelation*, is sometimes useful for diagnosing the form of a stationary time series model. This concept relies on a much broader conceptual framework called *linear prediction*, which closely connects to ideas from least-squares regression, and is discussed first:

Definition: Suppose (x_t) is stationary with $\mu = E[x_t] = 0$. The *best linear predictor* of

x_{n+m} is

$$x^n_{n+m} = \phi_{n1}x_n + \cdots + \phi_{nn}x_1,$$

in which $\phi_{n1}, \ldots, \phi_{nn}$ are the values of $\alpha_1, \ldots, \alpha_n$ that minimize

$$P^n_{n+m} \;=\; E[(x_{n+m} - x^n_{n+m})^2] = E[(x_{n+m} - \alpha_1 x_n - \cdots - \alpha_n x_1)^2].$$

The quantity x^n_{n+m} is sometimes called an m-step-ahead prediction, and P^n_{n+m} is called the expected squared prediction-error.

A formula for determining the coefficients of the best linear predictor is provided as follows.

Property: In the formula for x^n_{n+m}, the coefficients $\phi_{n1}, \ldots, \phi_{nn}$, satisfy

$$\gamma(m+k-1) \;=\; \phi_{n1}\gamma(k-1) + \cdots + \phi_{nn}\gamma(k-n)$$

for $k = 1, \ldots, n$, where $\gamma(h)$ is the autocovariance function. Dividing each side of this formula by $\gamma(0)$ yields an equivalent formula, wherein the autocorrelation function replaces the autocovariance function,

$$\rho(m+k-1) \;=\; \phi_{n1}\rho(k-1) + \cdots + \phi_{nn}\rho(k-n).$$

Proof: Use calculus to deduce

$$-2E[(x_{n+m} - \phi_{n1}x_n - \cdots - \phi_{nn}x_1)x_{n-k+1}] \;=\; 0,$$

hence

$$E[x_{n+m}x_{n-k+1}] \;=\; \phi_{n1}E[x_n x_{n-k+1}] + \cdots + \phi_{nn}E[x_1 x_{n-k+1}],$$

for $k = 1, \ldots n$. Stationarity and $\mu = 0$ imply

$$E[x_s x_t] = Cov[x_s, x_t] = \gamma(s-t),$$

which translates the above to

$$\gamma(m+k-1) \;=\; \phi_{n1}\gamma(k-1) + \cdots + \phi_{nn}\gamma(k-n)$$

□

As a simple example of how to use the linear prediction formula, suppose (x_t) is $AR(1)$, $n = 1$, and $m = 1$; then,

$$x^1_2 \;=\; \phi_{11}x_1$$

and

$$\gamma(1) \;=\; \phi_{11}\gamma(0).$$

Solving, we have that

$$\phi_{11} \quad = \quad \gamma(1)/\gamma(0) = \rho(1) = \phi_1.$$

The final step of this example requires alludes to a result that we will deduce in the next learning unit, which is that the autocorrelation function of an $AR(1)$ time series is $\rho(h) = \phi_1^h$.

In general, the above property identifies that the coefficients $\phi_{n1}, \ldots, \phi_{nn}$ are the solutions to a system of linear equations, and so might be calculated using general methods of matrix algebra. This is carefully discussed in the next section, below. Alternatively, specialized recursive computational algorithms have been developed for complex or other situations where general methods become burdensome. For the case $m = 1$, the following example describes one such algorithm, which calculates both the coefficients $\phi_{n1}, \ldots, \phi_{nn}$ and the expected squared prediction-errors P_{n+1}^n.

Example: The Durbin-Levinson algorithm

As stated, above, a general solution for finding the coefficients, $\phi_{n1}, \ldots, \phi_{nn}$, of the best linear predictor is

$$x_{n+m}^n = \phi_{n1} x_n + \cdots + \phi_{nn} x_1,$$

which is found by solving

$$\gamma(m + k - 1) \quad = \quad \phi_{n1}\gamma(k - n) + \cdots + \phi_{nn}\gamma(k - 1)$$

across $k = 1, \ldots, n$.

The Durbin-Levinson algorithm is an efficient, iterative algorithm, which implements the task of calculating the coefficients, $\phi_{n1}, \ldots, \phi_{nn}$ as follows. To start, initialize the algorithm by setting $\phi_{00} = 0$ and $P_1^0 = \gamma(0)$. Subsequently, recall the autocorrelation function, $\rho(h)$, and calculate iteratively as $n = 1, 2, \ldots$

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)} \quad \text{and} \quad P_{n+1}^n = (1 - \phi_{nn}^2) P_n^{n-1},$$

along the way calculating

$$\phi_{nk} \quad = \quad \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}$$

for $k = 1, 2, \ldots, n-1$, in steps such that $n \geq 2$.

As an example of how this algorithm executes, observe that, for $n = 1$,

$$\phi_{11} = \rho(1) \quad \text{and} \quad P_2^1 = \gamma(0)(1 - \phi_{11}^2);$$

for $n = 2$,

$$\phi_{22} = \frac{\rho(2) - \phi_{11}\rho(1)}{1 - \phi_{11}\rho(1)}, \ \phi_{21} = \phi_{11} - \phi_{22}\phi_{11}, \ \text{and} \ P_3^2 = \gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2);$$

and, for $n = 3$,

$$\phi_{33} = \frac{\rho(3) - \phi_{21}\rho(2) - \phi_{22}\rho(1)}{1 - \phi_{21}\rho(1) - \phi_{22}\rho(2)},$$

$$\phi_{32} = \phi_{22} - \phi_{33}\phi_{21},\ \phi_{31} = \phi_{21} - \phi_{33}\phi_{22},$$

$$\text{and } P_4^3 = \gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2)(1 - \phi_{33}^2).$$

□

Our tools are now in place to define partial autocorrelation.

Definition: Suppose (x_t) is a stationary Gaussian time series. Fix some integer $h > 1$, and define the quantities

$$\begin{aligned} \hat{x}_{t+h} &= a_1 x_{t+h-1} + \cdots + a_{h-1} x_{t+1} \\ \hat{x}_t &= b_1 x_{t+1} + \cdots + b_{h-1} x_{t+h-1} \end{aligned}$$

for coefficients $a_1, \ldots, a_{h-1}$ and $b_1, \ldots, b_{h-1}$ that respectively minimize the quantities

$$E[(x_{t+h} - \hat{x}_{t+h})^2] \quad \text{and} \quad E[(x_t - \hat{x}_t)^2].$$

The *partial autocorrelation function (PACF)* at lag h is

$$\phi_{hh} = Corr[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t].$$

At $h = 1$, the PACF is $\phi_{11} = \rho(1)$.

In the definition, $\hat{x}_{t+h}$ is said to be the *fitted value* in the linear regression of x_{t+h} on $(x_{t+h-1}, \cdots, x_{t+1})$, and $\hat{x}_t$ is the fitted value in the linear regression of x_t on $(x_{t+1}, \cdots, x_{t+h-1})$. Another way to see this is that we are contemplating the prediction of x_{t+h} and x_t from the values between them, $x_{t+h-1}, \cdots, x_{t+1}$. The connection of these ideas to linear prediction provides the following.

Property: As defined, above, the quantities $\hat{x}_{t+h}$ and $\hat{x}_t$ are such that the coefficients $a_1, \ldots, a_{h-1}$ and $b_1, \ldots, b_{h-1}$ satisfy $a_k = b_k$ for $k = 1, \ldots, h-1$, and

$$\rho(k) = a_1\rho(k-1) + \cdots + a_{h-1}\rho(k-h+1).$$

Proof: The equations that determine $a_1, \ldots, a_{h-1}$ are deduced from their mathematical parallel with the equations that determine the best linear predictor x_h^{h-1}. The derive the equations that determine $b_1, \ldots, b_{h-1}$, use calculus to deduce

$$-2E[(x_t - b_1 x_{t+1} - \cdots - b_{h-1} x_{t+h-1})x_{t+k}] = 0,$$

hence

$$E[x_t x_{t+k}] = b_1 E[x_{t+1}x_{t+k}] + \cdots + b_{h-1}E[x_{t+h-1}x_{t+k}],$$

from which it follows that

$$\gamma(k) \;=\; b_1\gamma(k-1) + \cdots + b_{h-1}\gamma(k-h+1).$$

Divide each side by $\gamma(0)$ to achieve the desired result. □

An important implication of the linear prediction setup is that the coefficient ϕ_{hh} in the one-step-ahead prediction formula

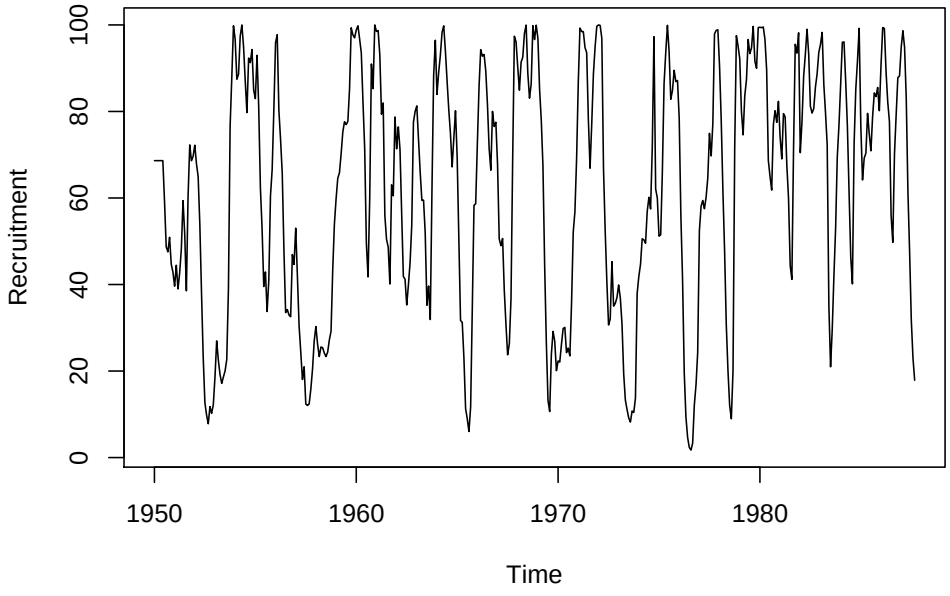
$$x_{h+1}^h = \phi_{h1}x_h + \cdots + \phi_{hh}x_1,$$

identifies with the value ϕ_{hh} of the partial autocorrelation function. This connection is carefully explained in the next section of these notes. In any case, the above property implies that the Durbin-Levinson algorithm admits an efficient method for calculating the PACF.

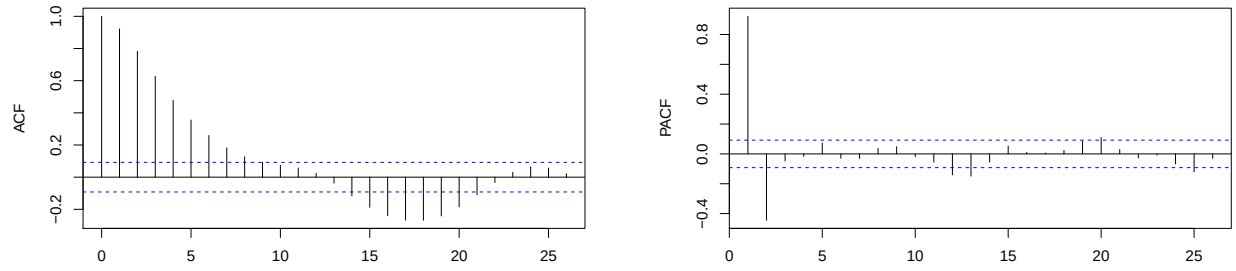
As will be explored in a subsequent learning unit, a central motivation for the PACF is its potential use in identifying a time-series model, particularly when considering an autoregressive model. The following example offers a demonstration of the *sample* PACF, which is defined as the PACF that is calculated upon substituting the sample ACF, $\hat{\gamma}(h)$, for the ACF, $\gamma(h)$, in the Durbin-Levinson algorithm.

Example: Fish population recruitment

The time series plotted in the figure below is of monthly fish population counts in the Pacific Ocean over the period 1950-1988.



The sample ACF and PACF calculated on the fish population are shown below.



Notice that both the sample ACF and PACF are plotted with cutoff values for assessing whether the plotted quantity is different from zero. Recall the cutoff value of the ACF is $2/\sqrt{n}$. That of the PACF arrises from a different formula.

A clear pattern is seen in the plots by which the ACF tapers while the PACF appears to cut off suddenly at lag $h = 2$; that is $\phi_{hh} \neq 0$ for $h \leq 2$ and $\phi_{hh} = 0$ for $h > 2$. As will be discussed, this is characteristic of an $AR(2)$ model. Further analysis might proceed under the assumption of that form of model, and subsequently move toward estimating the autoregressive parameters ϕ_1 and ϕ_2. □

Section 2B.4: Matrix formulas and mathematical connections.

Summarizing some of the core mathematical results of the last section's discussion, the partial autocorrelation at lag h of a mean-zero stationary time series (x_t) is

$$\alpha_{hh} = Corr[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t],$$

where

$$\begin{aligned} \hat{x}_{t+h} &= a_1 x_{t+h-1} + \cdots + a_{h-1} x_{t+1} \\ \hat{x}_t &= a_1 x_{t+1} + \cdots + a_{h-1} x_{t+h-1} \end{aligned}$$

are fitted values, whose coefficients $a_1, \ldots, a_{h-1}$ satisfy

$$\rho(j) = \sum_{k=1}^{h-1} a_k \rho(j-k)$$

for $j = 1, \ldots, h-1$. The latter formula identifies the $a_1, \ldots, a_{h-1}$ that define $\hat{x}_{t+h}$ and $\hat{x}_t$ as best linear predictors. An equivalent formula that identifies the $a_1, \ldots, a_{h-1}$ is

$$\gamma(j) = \sum_{k=1}^{h-1} a_k \gamma(j-k),$$

for $j = 1, \ldots, h-1$, which replaces the autocorrelation function, $\rho(h)$, in the previous formula with the autocovariance function, $\gamma(h)$.

In the above summary, notice that the previous notation ϕ_{hh} is replaced with α_{hh}. This is done to distinguish the definition of partial correlation just given from its characterization in terms of linear prediction, where ϕ_{hh} is a coefficient produced in the Durbin-Levinson algorithm. An objective of this section is to establish that $\phi_{hh} = \alpha_{hh}$, and so we start with distinct notation.

An explicit mathematical formula for partial autocorrelation implied from the above definition is derived as follows. First note that, since $h-j = 1, \ldots, h-1$ whenever $j = 1, \ldots, h-1$, resubscripting the formula $\gamma(j) = \sum_{k=1}^{h-1} a_k \gamma(j-k)$ implies

$$\gamma(h-j) = \sum_{k=1}^{h-1} a_k \gamma(h-j-k),$$

for $j = 1, \ldots, h-1$, from which, upon writing $\hat{x}_{t+h} = \sum_{k=1}^{h-1} a_k x_{t+h-k}$, it follows that

$$E[\hat{x}_{t+h} x_{t+j}] = \sum_{k=1}^{h-1} a_k E[x_{t+h-k}, x_{t+j}] = \sum_{k=1}^{h-1} a_k \gamma(h-j-k) = \gamma(h-j).$$

Noting that $E[x_{t+h} x_{t+j}] = \gamma(h-j)$, it has thus been deduced that $E[(x_{t+h} - \hat{x}_{t+h}) x_{t+j}] = 0$ for $j = 1, \ldots, h-1$. Because $E[x_t] = 0$, this formula equivalent to

$$Cov[x_{t+h} - \hat{x}_{t+h}, x_{t+j}] = 0,$$

for $j = 1, \ldots, h-1$.

We can now derive the building blocks of the general partial autocorrelation formula. For this, we use the property that $\hat{x}_t$ written in summation notation provides two equivalent expressions, $\hat{x}_t = \sum_{j=1}^{h-1} a_j x_{t+j} = \sum_{j=1}^{h-1} a_j x_{t+h-j}$, whose equivalence is established by resubscripting under the property $h - j = 1, \ldots, h-1$ whenever $j = 1, \ldots, h-1$. Now, apply $\hat{x}_t = \sum_{j=1}^{h-1} a_j x_{t+j}$, the linearity of covariance, and the above zero-covariance property, to observe

$$Cov[x_{t+h} - \hat{x}_{t+h}, \hat{x}_t] = \sum_{j=1}^{h-1} a_j Cov[x_{t+h} - \hat{x}_{t+h}, x_{t+j}] = 0.$$

Subsequently, apply $\hat{x}_t = \sum_{j=1}^{h-1} a_j x_{t+h-j}$ to see that

$$\begin{aligned}
Cov[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t] &= Cov[x_{t+h} - \hat{x}_{t+h}, x_t] - Cov[x_{t+h} - \hat{x}_{t+h}, \hat{x}_t] \\
&= Cov[x_{t+h} - \hat{x}_{t+h}, x_t] \\
&= Cov[x_{t+h}, x_t] - Cov[\hat{x}_{t+h}, x_t] \\
&= Cov[x_{t+h}, x_t] - \sum_{j=1}^{h-1} a_j Cov[x_{t+h-j}, x_t] \\
&= \gamma(h) - \sum_{j=1}^{h-1} a_j \gamma(h-j).
\end{aligned}$$

A similar deduction provides

$$\begin{aligned} Var[x_{t+h} - \hat{x}_{t+h}] &= Cov[x_{t+h} - \hat{x}_{t+h}, x_{t+h} - \hat{x}_{t+h}] \\ &= Cov[x_{t+h} - \hat{x}_{t+h}, x_{t+h}] \\ &= \gamma(0) - \sum_{j=1}^{h-1} a_j \gamma(-j) \end{aligned}$$

and

$$\begin{aligned} Var[x_t - \hat{x}_t] &= Cov[x_t - \hat{x}_t, x_t - \hat{x}_t] \\ &= Cov[x_t - \hat{x}_t, x_t] \\ &= \gamma(0) - \sum_{j=1}^{h-1} a_j \gamma(j). \end{aligned}$$

Putting these pieces together provides the sought-after partial autocorrelation formula:

$$\begin{aligned} \alpha_{hh} &= \frac{Corr[x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t]}{\sqrt{Var[x_{t+h} - \hat{x}_{t+h}] Var[x_t - \hat{x}_t]}} \\ &= \frac{\gamma(h) - \sum_{j=1}^{h-1} a_j \gamma(h-j)}{\gamma(0) - \sum_{j=1}^{h-1} a_j \gamma(j)} \\ &= \frac{\rho(h) - \sum_{j=1}^{h-1} a_j \rho(h-j)}{1 - \sum_{j=1}^{h-1} a_j \rho(j)}, \end{aligned}$$

wherein the last step is made by dividing the numerator and denominator by $\gamma(0)$.

A deeper exploration of this partial autocorrelation formula arises by expressing it in vector-matrix notation. To start, let us rewrite the fitted-value coefficients as follows: The criterion specifying the $a_1, \ldots, a_{h-1}$ may be rewritten $\boldsymbol{\rho}_{h-1} = \boldsymbol{R}_{h-1}\boldsymbol{a}$, where

$$\boldsymbol{\rho}_{h-1} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h-1) \end{bmatrix}, \ \boldsymbol{R}_{h-1} = \begin{bmatrix} \rho(0) & \rho(-1) & \cdots & \rho(-h+2) \\ \rho(1) & \rho(0) & \cdots & \rho(-h+3) \\ \vdots & & & \\ \rho(h-2) & \rho(h-3) & \cdots & \rho(0) \end{bmatrix}, \text{ and } \boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{h-1} \end{bmatrix}.$$

The coefficients themselves are thus given by the solution $\boldsymbol{a} = \boldsymbol{R}_{h-1}^{-1}\boldsymbol{\rho}_{h-1}$, where $\boldsymbol{R}_{h-1}^{-1}$ is the inverse-matrix of $\boldsymbol{R}_{h-1}$. By additionally defining the vector

$$\tilde{\boldsymbol{\rho}}_{h-1} = \begin{bmatrix} \rho(h-1) \\ \rho(h-2) \\ \vdots \\ \rho(1) \end{bmatrix},$$

the formula for partial autocorrelation becomes

$$\alpha_{hh} = \frac{\rho(h) - \boldsymbol{a}^T \tilde{\boldsymbol{\rho}}_{h-1}}{1 - \boldsymbol{a}^T \boldsymbol{\rho}_{h-1}} = \frac{\rho(h) - \boldsymbol{\rho}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}}{1 - \boldsymbol{\rho}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \boldsymbol{\rho}_{h-1}}.$$

Let us go further, aiming to use this expression to connect partial autocorrelation to the Durbin-Levinson algorithm. A step in making this connection is to show that the term $\boldsymbol{\rho}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \boldsymbol{\rho}_{h-1}$ in the partial autocorrelation formula may be rewritten $\tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}$. To see this, let us contemplate the action of a square matrix of the form

$$\boldsymbol{B} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

A little though about this matrix will convince you that if $\boldsymbol{C}$ is any other matrix of the same dimension, then $\boldsymbol{BC}$ is $\boldsymbol{C}$ with its rows reversed while $\boldsymbol{CB}$ is $\boldsymbol{C}$ with its columns reversed. Now, the structure of the matrix $\boldsymbol{R}_{h-1}$ is such that reversing its rows is identical to reversing its columns; therefore, $\boldsymbol{B}\boldsymbol{R}_{h-1} = \boldsymbol{R}_{h-1}\boldsymbol{B}$.

Next, observe that $\tilde{\boldsymbol{\rho}}_{h-1} = \boldsymbol{B}\boldsymbol{\rho}_{h-1}$, and since $\boldsymbol{\rho}_{h-1} = \boldsymbol{R}_{h-1}\boldsymbol{a}$ we have

$$\tilde{\boldsymbol{\rho}}_{h-1} = \boldsymbol{B}\boldsymbol{R}_{h-1}\boldsymbol{a} = \boldsymbol{R}_{h-1}\boldsymbol{B}\boldsymbol{a} = \boldsymbol{R}_{h-1}\tilde{\boldsymbol{a}},$$

where

$$\tilde{\boldsymbol{a}} = \begin{bmatrix} a_{h-1} \\ a_{h-2} \\ \vdots \\ a_1 \end{bmatrix}.$$

This implies $\tilde{\boldsymbol{a}} = \boldsymbol{R}_{h-1}^{-1}\tilde{\boldsymbol{\rho}}_{h-1}$. Putting these pieces together, we see that

$$\boldsymbol{\rho}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \boldsymbol{\rho}_{h-1} = \boldsymbol{a}^T \boldsymbol{\rho}_{h-1} = \sum_{j=1}^{h-1} a_j \rho(j) = \sum_{j=1}^{h-1} a_{h-j+1} \rho(h-j+1) = \tilde{\boldsymbol{a}}^T \tilde{\boldsymbol{\rho}}_{h-1} = \tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}.$$

Our formula is now

$$\alpha_{hh} = \frac{\rho(h) - \boldsymbol{a}^T \tilde{\boldsymbol{\rho}}_{h-1}}{1 - \boldsymbol{a}^T \boldsymbol{\rho}_{h-1}} = \frac{\rho(h) - \boldsymbol{\rho}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}}{1 - \tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}}.$$

We are very close to connecting connecting the partial autocorrelation formula to the Durbin-Levinson algorithm. Observe that the best linear predictor of x_{h+1} from $x_1, \ldots, x_h$ is

$$x_{h+1}^h = \phi_{h1} x_h + \cdots + \phi_{hh} x_1,$$

where

$$\rho(k) = \sum_{j=1}^{h} \phi_{hj}\rho(k-j),$$

for $k = 1, \ldots, h$. Writing in vector-matrix notation, the criterion specifying the $\phi_{h1}, \ldots, \phi_{hh}$ is $\boldsymbol{\rho}_h = \boldsymbol{R}_h\boldsymbol{\phi}$, where

$$\boldsymbol{\rho}_h = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{bmatrix}, \ \boldsymbol{R}_h = \begin{bmatrix} \rho(0) & \rho(-1) & \cdots & \rho(-h+1) \\ \rho(1) & \rho(0) & \cdots & \rho(-h+2) \\ \vdots & & & \\ \rho(h-1) & \rho(h-2) & \cdots & \rho(0) \end{bmatrix}, \text{ and } \boldsymbol{\phi} = \begin{bmatrix} \phi_{h1} \\ \phi_{h2} \\ \vdots \\ \phi_{hh} \end{bmatrix}.$$

By partitioning the above vectors and matrices, another way to write the criterion $\boldsymbol{\rho}_h = \boldsymbol{R}_h\boldsymbol{\phi}$ is

$$\begin{bmatrix} \boldsymbol{\rho}_{h-1} \\ \rho(h) \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{h-1} & \tilde{\boldsymbol{\rho}}_{h-1} \\ \tilde{\boldsymbol{\rho}}_{h-1}^T & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_1 \\ \phi_{hh} \end{bmatrix}, \text{ where } \boldsymbol{\phi}_1 = \begin{bmatrix} \phi_{h1} \\ \vdots \\ \phi_{h,h-1} \end{bmatrix}.$$

Solving this formula provides

$$\begin{bmatrix} \boldsymbol{\phi}_1 \\ \phi_{hh} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{h-1} & \tilde{\boldsymbol{\rho}}_{h-1} \\ \tilde{\boldsymbol{\rho}}_{h-1}^T & 1 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\rho}_{h-1} \\ \rho(h) \end{bmatrix}.$$

We are at a point where we need additional formula from vector-matrix arithmetic: Suppose the matrix $\boldsymbol{A}$ is partitioned according to

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}.$$

The general formula for the inverse of a partitioned matrix provides

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{B}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}^{-1} \\ -\boldsymbol{B}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{B}_{22}^{-1} \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{B}_{11} &= \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} \\ \boldsymbol{B}_{22} &= \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}. \end{aligned}$$

Applying the general inverse formula to the problem at hand, equating $\boldsymbol{A}_{11} = \boldsymbol{R}_{h-1}$, $\boldsymbol{A}_{12} = \tilde{\boldsymbol{\rho}}_{h-1}$, $\boldsymbol{A}_{21} = \tilde{\boldsymbol{\rho}}_{h-1}^T$, and $\boldsymbol{A}_{22} = 1$, we see that

$$\phi_{hh} = -\boldsymbol{B}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{\rho}_{h-1} + \boldsymbol{B}_{22}^{-1}\rho(h) = -\boldsymbol{B}_{22}^{-1}\tilde{\boldsymbol{\rho}}_{h-1}^T\boldsymbol{R}_{h-1}^{-1}\boldsymbol{\rho}_{h-1} + \boldsymbol{B}_{22}^{-1}\rho(h),$$

where

$$\boldsymbol{B}_{22} = 1 - \tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}.$$

Re-expressed, the relevant formula is

$$\phi_{hh} = \frac{\rho(h) - \tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \boldsymbol{\rho}_{h-1}}{1 - \tilde{\boldsymbol{\rho}}_{h-1}^T \boldsymbol{R}_{h-1}^{-1} \tilde{\boldsymbol{\rho}}_{h-1}},$$

which is identical to the formula for α_{hh} derived from the definition of partial autocorelation. It has therefore been deduced that $\phi_{hh} = \alpha_{hh}$, hence the Durbin-Levinson theorem may be used to calculate the partial autocorrelation function.