The partial autocorrelation function

Intuitively, in a $\mathrm{MA}(q)$ process the ACF will be zero for lags greater than q. Moreover, because $\theta_q \neq 0$, the ACF will not be zero at lag q. Thus, the ACF provides a considerable amount of information about the order of the dependence when the process is a moving average process.

If the process, however, is ARMA or AR, the ACF alone tells us little about the orders of dependence. Hence, it is worthwhile pursuing a function that will behave like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF).

Definition

If X, Y and Z are random variables, then the partial correlation between X and Y given Z is obtained by regressing X on Z to obtain $\hat{X}$, ressing Y on Z to obtain $\hat{Y}$, and then calculating

$$\rho_{\mathrm{XY}|Z} = \mathrm{corr}\{X - \hat{X}, Y - \hat{Y}\}$$

The idea is that $\rho_{\mathrm{XY}|Z}$ measures the correlation between X and Y with the linear effect of Z removed (or partialled out).

Motivation

To motivate the idea for time series, consider a causal $\mathrm{AR}(1)$ model, $x_t = \phi x_{t-1} + w_t$. Then,

$$\begin{aligned}\gamma_x(2) = \mathrm{cov}(x_t, x_{t-2}) &= \mathrm{cov}(\phi x_{t-1} + w_t, x_{t-2}) \\ &= \mathrm{cov}(\phi^2 x_{t-2} + \phi w_{t-1} + w_t, x_{t-2}) = \phi^2 \gamma_x(0)\end{aligned}$$

This result follows from causality because x_{t-2} involves $\{w_{t-2}, w_{t-3}, \ldots\}$, which are all uncorrelated with w_t and w_{t-1}. The correlation between x_t and x_{t-2} is not zero, as it would be for an $\mathrm{MA}(1)$, because x_t is dependent on x_{t-2} through x_{t-1}.

Suppose we break this chain of dependence by removing (or partial out) the effect x_{t-1} . That is, we consider the correlation between $x_t - \phi x_{t-1}$ and $x_{t-2} - \phi x_{t-1}$, because it is the correlation between x_t and x_{t-2} with the linear dependence of each on x_{t-1} removed.

$$\mathrm{cov}(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = \mathrm{cov}(w_t, x_t - \phi x_{t-1}) = 0$$

Linear dependency

To formally define the PACF for mean-zero stationary time series, let $\hat{x}_{t+h}$, for $h \geq 2$, denote the regression of x_{t+h} on $\{x_{t+h-1}, x_{t+h-2}, \ldots, x_{t+1}\}$, which we write as

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \cdots + \beta_{h-1} x_{t+1}$$

No intercept term is needed because the mean of x_t is zero. In addition, let $\hat{x}_t$ denote the regression of x_t on $\{x_{t+1}, x_{t+2}, \ldots, x_{t+h-1}\}$, then

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h-1}$$

Because of stationarity, the coefficients, $\beta_1, \ldots, \beta_{h-1}$ are the same in both equations.

The PACF

Definición 1. *The partial autocorrelation function (PACF) of a stationary process, x_t, denoted ϕ_{hh}, for $h = 1, 2, \ldots,$ is*

$$\phi_{11} = \text{corr}(x_{t+1}, x_t) = \rho(1)$$

and

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), \quad h \geq 2$$

The PACF, ϕ_{hh}, is the correlation between x_{t+h} and x_t with the linear dependence of $\{x_{t+1}, \ldots, x_{t+h-1}\}$ on each, removed.

Visual results

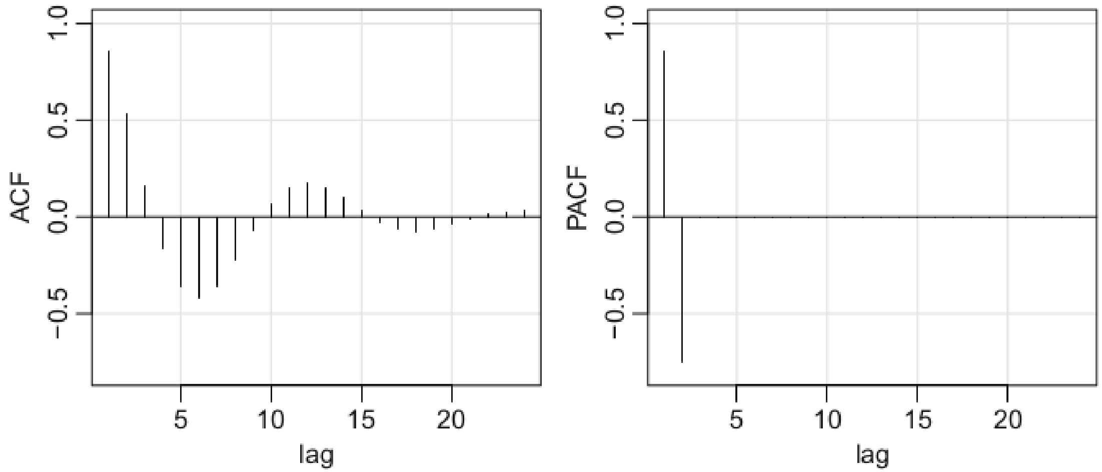


Fig. 3.5. The ACF and PACF of an AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -.75$

Behavior of the ACF and PACF for ARMA

	$\mathrm{AR}(p)$	$\mathrm{MA}(q)$	$\mathrm{ARMA}(p,q)$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

The PACF for MA models behaves much like the ACF for AR models. Also, the PACF for AR models behaves much like the ACF for MA models. Because an invertible ARMA model has an infinite AR representation, the PACF will not cut off.

Forecasting

In forecasting, the goal is to predict future values of a time series, x_{n+m}, $m = 1, 2, \ldots$, based on the data collected to the present, $x_{1:n} = \{x_1, x_2, \ldots, x_n\}$.

The minimum mean squared error predictor of x_{n+m} is

$$x_{n+m}^n = \mathbb{E}[x_{n+m} | x_{1:n}]$$

because the conditional expectation minimizes the mean squred error

$$\mathbb{E}[(x_{n+m} - g(x_{1:n}))^2]$$

Consider the one-step-ahead prediction:

$$x_{n+1}^{n} = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \cdots + \phi_{nn} x_1$$

By minimizing the expected values via least squares we get the solution

$$\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n$$

where $\Gamma_n = \{\gamma(k-j)\}_{j,k=1}^{n}$ is an $n \times n$ matrix, $\boldsymbol{\phi}_n = (\phi_{n1}, \ldots, \phi_{nn})^T$ is an $n \times 1$ vector, and $\boldsymbol{\gamma}_n = (\gamma(1), \ldots, \gamma(n))^T$ is an $x \times 1$ vector.

It is sometimes convenient to write the one-step-ahead forecast in vector notation

$$x_{n+1}^n = \boldsymbol{\phi}^T \boldsymbol{x}$$

where $\boldsymbol{x} = (x_n, x_{n-1}, \ldots, x_1)^T$.

The mean square one-step-ahead prediction error is

$$\begin{aligned}
\sigma_{n+1}^2 = P_{n+1}^n = \mathbb{E}[(x_{n+1} - x_{n+1}^n)^2] &= \mathbb{E}[(x_{n+1} - \boldsymbol{\phi}^T \boldsymbol{x})^2] = \mathbb{E}[(x_{n+1} - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{x})^2] \\
&= \mathbb{E}[x_{n+1}^2 - 2\boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{x}\, x_{n+1} + \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{x}\, \boldsymbol{x}^T \Gamma_n^{-1} \boldsymbol{\gamma}_n] \\
&= \gamma(0) - 2\boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n + \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \boldsymbol{\gamma}_n \\
&= \gamma(0) - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n
\end{aligned}$$

Equations $\boldsymbol{\phi}_n = \Gamma_n^{-1}\boldsymbol{\gamma}_n$ and $P_{n+1}^n = \gamma(0) - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n$ can be solved iteratively as

$$\phi_{00} = 0, \quad P_1^0 = \gamma(0)$$

For $n \geq 1$,

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

where, for $n \geq 2$

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1$$

The Innovations algorithm

The one-step-ahead predictors, x_{t+1}^t, and their mean-squared errors, P_{t+1}^t, can be calculated iteratively as

$$x_1^0 = 0, \quad P_1^0 = \gamma(0)$$

$$x_{t+1}^t = \sum_{j=1}^{t} \theta_{tj}(x_{t+1-j} - x_{t+1-j}^{t-j}), \quad t = 1, 2, \ldots$$

$$P_{t+1}^t = \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 P_{j+1}^j, \quad t = 1, 2, \ldots$$

where, for $j = 0, 1, \ldots, t-1$,

$$\theta_{t,t-j} = \left(\gamma(t-j) - \sum_{k=0}^{j-1} \theta_{j,j-k}\theta_{t,t-k} P_{k+1}^k \right) / P_{j+1}^j$$

Recall we can write the $\mathrm{ARMA}(p, q)$ model as its operator form as

$$\phi(B)x_t = \phi(B)\psi(B)w_t = \theta(B)w_t$$

Where (by solving the difference equation)

$$\psi_0 = 1$$

$$\psi_j = \theta_j + \sum_{k=1}^{\min(j,p)} \phi_k \psi_{j-k}$$

Where $\theta_j = 0$ for $j > q$, and $\phi_k = 0$ for $k > p$.

ARMA estimating parameters

Truncating at $p+q$, use the innovation algorithm to estimate $\psi_1, \ldots, \psi_{p+q}$ in

$$x_t = \psi_0 + \psi_1 w_{t-1} + \cdots + \psi_{p+q} w_{t-p-q}$$

Then for $k = q+1, \ldots, q+p$ solve for ϕ_i (using Durbin-Levinson algorithm)

$$\psi_k = \phi_1 \psi_{k-1} + \phi_2 \psi_{k-2} + \ldots + \phi_p \psi_{k-p}$$

Finally, use the estimated ϕ_i and ψ_i to estimate θ_i, for $i = 1, \ldots, q$ usings

$$\psi_j = \theta_j + \sum_{k=1}^{\min(j,p)} \phi_k \psi_{j-k}$$

Integrated Models for Nonstationary Data

In many situations, time series can be thought of as being composed of two components, a nonstationary trend component and a zero-mean stationary component.

For example, if we consider the model

$$x_t = \mu_t + y_t$$

where $\mu_t = \beta_0 + \beta_1 t$ and y_t is stationary. Differencing such a process will lead to a stationary process

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t$$

Another model that leads to first differencing is the case in which μ_t is stochastic and slowly varying according to a random walk. That is, $\mu_t = \mu_{t-1} + v_t$ where v_t is stationary.

Order d differencing

If μ_t in $x_t = \mu_t + y_t$ is a k-th order polynomial, $\mu_t = \sum_{j=0}^{k} \beta_j t_j$, then the differenced series $\nabla^k x_t$ is stationary. Stochastic trend models can also lead to higher order differencing. For example, suppose

$$\mu_t = \mu_{t-1} + v_t \quad \text{and} \quad v_t = v_{t-1} + e_t$$

where e_t is stationary. Then $\nabla x_t = v_t + \nabla y_t$ is not stationary, but

$$\nabla^2 x_t = e_t + \nabla^2 y_t$$

is stationary.

The integrated ARMA, or ARIMA, model is a broadening of the class of ARMA models to include differencing.

Integrated ARMA

Definición 2. *A process* x_t *is said to be* $\mathrm{ARIMA}(p,d,q)$ *if*

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

is $\mathrm{ARMA}(p,q)$. *In general, we will write the model as*

$$\phi(B)(1-B)^d x_t = \theta(B) w_t$$

If $\mathbb{E}[\nabla^d x_t] = \mu$, *we write the model as*

$$\phi(B)(1-B)^d x_t = \delta + \theta(B) w_t$$

where $\delta = \mu(1-\phi_1-\cdots-\phi_p)$

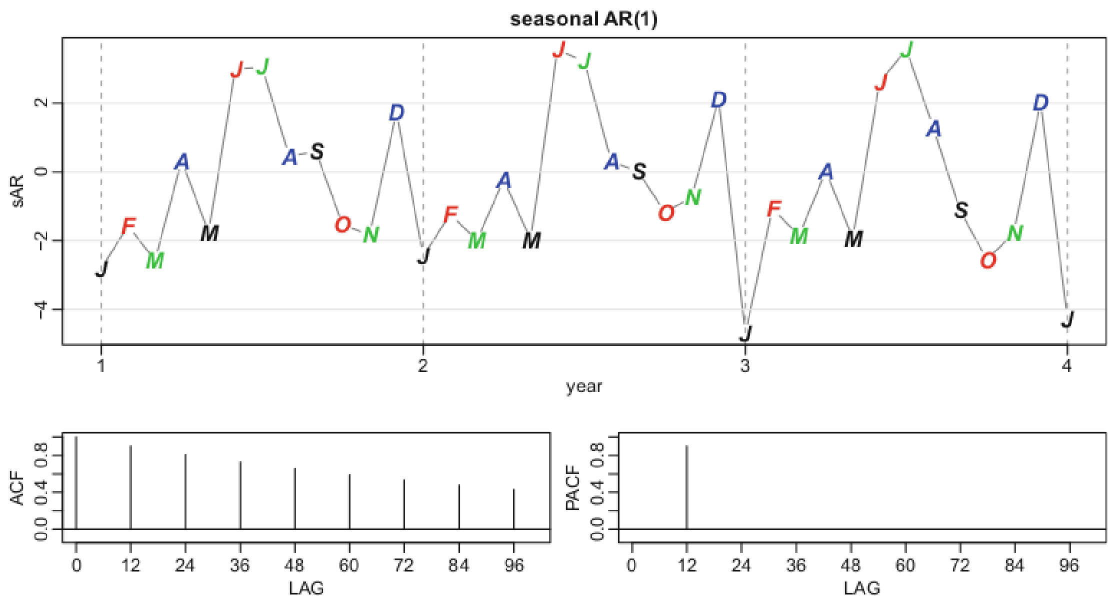
Building ARIMA Models

There are a few basic steps to fitting ARIMA models to time series data. These steps involve:

- plotting the data (inspect for anomalies)
- possibly transforming the data (stabilize variance, identify differencing d order),
- identifying the dependence orders of the model (Using ACF and PACF to identify q and p),
- parameter estimation (apply algorithms),
- diagnostics, and (validate)
- model choice.

Often, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag s.

For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of $s = 12$, because of the strong connections of all activity to the calendar year.



Seasonal operators

Because of this, it is appropriate to introduce autoregressive and moving average polynomials that identify with the seasonal lags. The resulting pure seasonal autoregressive moving average model, say, $\mathrm{ARMA}(P, Q)_s$, then takes the form

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t$$

where the operators

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{\mathrm{Ps}}$$

and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{\mathrm{Qs}}$$

are the *seasonal autoregressive operator* and the *seasonal moving average operator* of orders P and Q, respectively, with seasonal period s.

Seasonal ARMA

In general, we can combine the seasonal and nonseasonal operators into a multiplicative seasonal autoregressive moving average model, denoted by $\mathrm{ARMA}(p,q)\times(P,Q)_s$ and write

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(B)w_t$$

Also, we can remove seasonality via differencing. Then, a seasonal difference of order D is defined as:

$$\nabla_s^D x_t = (1-B^s)^D x_t$$

where $D=1,2,\ldots$, takes positive integer values. Typically, $D=1$ is sufficient to obtain seasonal stationarity.

Definición 3. *The multiplicative seasonal autoregressive integrated moving average model, or* SARIMA *model is given by*

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \Theta_Q(B^s)\theta(B)w_t$$

The general model is denoted as $\mathrm{ARIMA}(p, d, q) \times (P, D, Q)_s$.

How do you estimate the parameters?