Handout 9

Partial autocorrelation function (PACF)

Partial autocorrelations play an important role in time series analysis along with autocorrelations (ACF). The concept of partial autocorrelation is the same as that of partial correlation in regression. We will first review partial correlation in the context of Electricity bill for households (Handouts 2 and 3).

Electricity bill data (Handouts 2 and 3)

The electricity bill data has $n = 34$ households with $Y =$ monthly electricity bill, X_1=monthly income, $X_2 =$ number of persons in the household, X_3=living area.

The correlation matrix is:

	Bill	*Income*	*Persons*	*Area*
Bill	1	.837	.494	.905
Income		1	.143	.961
Person			1	.366
Area				1

For the purpose of discussion here, let us focus on the following two models

$$\hat{Y} = 358.4 + 0.0571X_1 + 55.09X_2 + 0.2811X_3, SSE = 550163, BIC = 343.6,$$
$$\hat{Y} = -574.41 + 0.2421X_1 + 85.34X_2, SSE = 578508, BIC = 341.8.$$

In Handout 3, you will find that the best model is regression of Y on X_2, X_3. However, we will focus here on the regression model of Y on X_1 and X_2. This model is quite reasonable since its BIC value is the second smallest. Note that $Corr(Y, X_3) = 0.905$. However, the selected model (i.e., regression of Y on X_1 and X_2) only has variables X_1 and X_2. We will see shortly that in the presence of X_1 and X_2, variable X_3 may not be needed for predicting Y. Is there a descriptive measure for this? One can use either coefficient of partial determination or partial correlation for this purpose.

When we fit a regression model for predicting Y from X_1 and X_2, let us denote the residual sum of squares by $SSE(X_1, X_2)$. If we fit Y on X_1, X_2 and X_3, then we will denote the residual sum of squares by $SSE(X_1, X_2, X_3)$. For the electric bill data, $SSE(X_1, X_2) = 578508, SSE(X_1, X_2, X_3) = 550163$.

The **coefficient of partial determination** is defined to be the proportional reduction in the residual sum of squares due to adding variable X_3 to the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, i.e.,

$$r^2_{YX_3|X_1X_2} = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{SSE(X_1, X_2)} = \frac{578508 - 550163}{578508} = 0.0490$$

For the electric bill data, there is only a 4.9% reduction in the residual sum of squares when we add variable X_3 to the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. This indicates that, in order to predict Y well, variable X_3 may not be needed if we already have variables X_1 and X_2 in the model.

Partial correlation $r_{YX_3|X_1X_2}$, between Y and X_3 given X_1 and X_2 is defined to be

$$r_{YX_3|X_1X_2} = \pm\sqrt{r^2_{YX_3|X_1X_2}},$$

with $'+'$ sign if $\hat{\beta}_3 > 0$ and $'-'$ sign if $\hat{\beta}_3 < 0$. Since $\hat{\beta}_3 > 0$ for the electric bill data, $r_{YX_3|X_1X_2} = \sqrt{0.0490} = 0.2214..$

It is important to point out that we may also calculate partial correlations $r_{YX_1|X_2X_3}$ or $r_{YX_2|X_1X_3}$ etc. There is an important mathematical result which provides an alternative interpretation of partial correlation.

Fact: Regress Y on X_1 and X_2, call the fitted values $\hat{Y}$. Regress X_3 on X_1 and X_2, call the fitted values $\hat{X}_3$. Then $r_{YX_3|X_1X_2} = Corr(Y - \hat{Y}, X_3 - \hat{X}_3)$.

[Similar result holds for $r_{YX_1|X_2X_3}$. Regress Y on X_2 and X_3, call the fitted values $\hat{Y}$, and regress X_1 on X_2 and X_3, and call the fitted values $\hat{X}_1$. Then $r_{YX_1|X_2X_3} = Corr(Y - \hat{Y}, X_1 - \hat{X}_1)$.]

Partial autocorrelation function (PACF)

(**Can be useful for identifying the order of AR models**)

We will begin with the definition.

Partial autocorrelation of order 2 ($PACF(2)$) is defined to be the partial correlation between X_t and X_{t-2}, given X_{t-1}.

Partial autocorrelation of order 3 ($PACF(3)$) is defined to be the partial correlation between X_t and X_{t-3}, given X_{t-1} and X_{t-2}.

In general, partial autocorrelation of order h ($PACF(h)$) is defined to be the partial correlation between X_t and X_{t-h} given $X_{t-1}, ..., X_{t-h+1}$.

By convention, $PACF(1)$ is defined to be $ACF(1) = Corr(X_t, X_{t-1})$.

Before we into further details, let us first write down an important result and how it is applied.

Fact: If the stationary sequence $\{X_t\}$ follows an $AR(p)$ model, then $PACF(h) = 0, h = p+1, p+2, ...$

PACF plot is like the ACF plot except that estimates of $PACF$ of lags 1,2,... are plotted along with $\pm 1.96/\sqrt{n}$ bars. For any lag, if the estimated PACF is inside the bars, that may be taken as an indication that the true PACF of that lag may be negligible.

Example 1: If X_t follows an $AR(1)$ model, then

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t, i.e., \ X_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t, with \ \beta_0 = (1-\phi_1)\mu, \ \beta_1 = \phi_1,$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2. Recall that correlation between X_t and X_{t-2} is $\rho(2) = \phi^2 \neq 0$ unless $\phi = 0$. However, a simple calculation will show that partial correlation between X_t and X_{t-2}, given X_{t-1} is zero. Why? When regressing X_t on X_{t-1}, i.e, trying to forecast X_t from X_{t-1}, the regression function is $X_t^{(f)} = \beta_0 + \beta_1 X_{t-1}$ [so $X_t^{(f)}$ is the fitted value when regressing X_t on X_{t-1}]. Note that $X_t - X_t^{(f)} = \varepsilon_t$. When we regress X_{t-2} on X_{t-1}, i.e., trying to backcast X_{t-2} from X_{t-1}, the fitted value $X_{t-2}^{(b)}$ is a linear function of X_{t-1}, and consequently $X_{t-2} - X_{t-2}^{(b)}$ depends only on X_{t-2} and X_{t-1}. Note that $X_t - X_t^{(f)} = \varepsilon_t$ is independent of $X_{t-1}, X_{t-2}, ...$ and hence of $X_{t-2} - X_{t-2}^{(b)}$. Hence the partial correlation between X_t and X_{t-2} given X_{t-1} is $Corr(X_t - X_t^{(f)}, X_{t-2} - X_{t-2}^{(b)}) = 0$.

How does one calculate PACF from the data? How does one justify using the $\pm 1.96/\sqrt{n}$ bar to check if the PACF of any particular lag is zero? Here is another technical fact.

Fact: a) If we fit an $AR(p+1)$ model, i.e., fit $\phi(B)(X_t - \mu) = \varepsilon_t$, where $\phi(B) = 1 - \phi_1 B - \cdot\cdot \cdot - \phi_{p+1}B^{p+1}$, then an estimate of $PACF(p+1)$ is given by the estimate $\hat{\phi}_{p+1}$ of ϕ_{p+1}.
b) If the series follows an $AR(p)$ model, then $0 = \phi_{p+1} = \phi_{p+2} = \cdot\cdot\cdot$. Then the distribution of $\hat{\phi}_h$ is approximately normal with mean 0 and variance $1/n$, for any $h = p+1, p+2, ...$

In general, partial autocorrelations do not have simple formulas. In the Appendix, there is a discussion on this issue. However, it is possible to find a formula for PACF of lag 2 and it is given in the following Excercise.

Exercise: Let $\{X_t\}$ be stationary (not necessarily $AR(p)$) with mean μ. Denote $\rho(1)$ by ϕ.
a) Show that the best linear predictor (forecast) of X_t from X_{t-1} is $X_t^{(f)} = \mu + \phi(X_{t-1} - \mu)$.
b) Show that the best predictor (backcast) of X_{t-2} from X_{t-1} is $X_{t-2}^{(b)} = \mu + \phi(X_{t-1} - \mu)$.
c) Show that $Var(X_t - X_t^{(f)}) = (1+\phi^2)\gamma(0) - 2\phi\gamma(1)$.
d) Show that $Var(X_{t-2} - X_{t-2}^{(f)}) = (1+\phi^2)\gamma(0) - 2\phi\gamma(1)$.
e) Show that $Cov(X_t - X_t^{(f)}, X_{t-2} - X_{t-2}^{(b)}) = \gamma(2) - 2\phi\gamma(1) + \phi^2\gamma(0)$.
f) Show that the partial correlation between X_t and X_{t-2} given X_{t-1} is given by $[\gamma(2) - 2\phi\gamma(1) + \phi^2\gamma(0)]/[(1+\phi^2)\gamma(0) - 2\phi\gamma(1)]$.

Diagnostic methods (ACF and PACF plots):

In time series analysis, both ACF and PACF plots are used. It is useful to keep in mind that for an $MA(q)$ model, autocorrelations of lag $q+1$ or higher are all zero. Whereas for an $AR(p)$ model, partial autocorrelations of order $p+1$ or higher are zero. So the ACF plot is useful for detecting moving average models and the PACF plot is useful in detecting autoregressive models. For instance, if all the autocorrelations of lag 3 or higher are all negligible, then an $MA(2)$ model may provide a reasonable description of the data. If the partial autocorrelations of lag 3 or higher are negligible, then $AR(2)$ may be a reasonable model for the data. The following table provides a summary.

Model	$AR(p)$	$MA(q)$	$ARMA(p,q)$
ACF	tails off	cuts off after lag q	tails off
PACF	cuts off after lag p	tails off	tails off

Recruitment series (estimated new fish):

This series has $n = 453$ observations from 1950-1987 (monthly). A plot is given below. The ACF plot suggests that the autocorrelations decay in a sinusoidal damped fashion: an issue to be discussed later. PACF plot shows that the partial autocorrelations of lag 3 or higher may be negligible. This suggests that an $AR(2)$ model may be reasonable in this case. Fitted AR(2) model is given below from R (using "sarima" function):

$\hat{\mu} = \bar{X} = 61.8585, \hat{\phi}_1 = 1.2512, \hat{\phi}_2 = -0.4612,$

$s(\hat{\mu}) = 4.0039, s(\hat{\phi}_1) = 0.0416, s(\hat{\phi}_2) = 0.0417.$

Plot the residuals from an $AR(2)$ fit is given below along with its ACF plot. [Residuals from an $AR(2)$ fit are: $\hat{\varepsilon}_t = X_t - \bar{X} - \{\hat{\phi}_1(X_{t-1} - \bar{X}) + \hat{\phi}_2(X_{t-2} - \bar{X})\}$]..

This ACF plot of residuals indicates that the residuals may be assumed to be white noise.

Tentative conclusion: an AR(2) model may be a reasonable description for the recruitment data.

In order to get a good predictive model, one should use a criterions such as AIC or AICC to select an appropriate model. However, the use of ACF and PACF plots may often lead us to a reasonable predictive model.

Appendix: Technical Issues

The following technical discussion on partial autocorrelation is true for any stationary series $\{X_t\}$, not just autoregressive models. You can check this easily for $p = 1$. If we try to predict X_t from the past p observations $X_{t-1}, ..., X_{t-p}$, then a linear predictor is of the form $c_0 + c_1 X_{t-1} + \cdots + c_p X_{t-p}$, where $c_0, ..., c_p$ are constants. Different choices of the constants $c's$ will lead to different linear predictors. Is there a best linear predictor? If so how does one characterize this? The answer is yes. If the mean of the process is μ, then the best linear predictor is of the form

$$\mu + \phi_p(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu),$$

where $\phi_1, ..., \phi_p$ are solutions of the following (theoretical Yule-Walker) equations

$$\begin{aligned}
\gamma(0)\phi_1 + \gamma(1)\phi_2 + \cdots + \gamma(p-1)\phi_p &= \gamma(1) \\
\gamma(1)\phi_1 + \gamma(0)\phi_2 + \cdots + \gamma(p-2)\phi_p &= \gamma(2) \\
&\vdots \\
\gamma(p-1)\phi_1 + \gamma(p-2)\phi_2 + \cdots + \gamma(0)\phi_p &= \gamma(p).
\end{aligned}$$

Employing vector-matrix notations, these equations are

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & & \\ \gamma(p-1) & \gamma(p-2) & & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix} \tag{1}$$

or

$$\Gamma_p \phi_p = \gamma_p,$$

where Γ_p is a $p \times p$ matrix whose element (i, j) is given by $\gamma(|i - j|)$, γ_p denotes the column vector with elements $\gamma(1), ..., \gamma(p)$ and ϕ_p denotes the column vector with elements $\phi_1, ..., \phi_p$.

In order to define partial correlation between X_t and X_{t-p-1} (i.e., PACF of lag $p + 1$), we need to predict X_t from $X_{t-1}, ..., X_{t-p}$ and get the residual (innovation). We also need to predict X_{t-p-1} from $X_{t-1}, ..., X_{t-p}$. In order to distingusigh these two predictions, we call the first forecasting and the second one backcasting. The best linear predictor (forecast) of X_t using $X_{t-1}, ..., X_{t-p}$ is

$$X_t^{(f)} = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \cdots + \phi_p(X_{t-p} - \mu),$$

where $\phi_1, ..., \phi_p$ are solutions of the equations given in (1).

Surprisingly, it turns out that the best predictor (backcast) of X_{t-p-1} using $X_{t-1}, ..., X_{t-p}$ is

$$X_{t-p-1}^{(b)} = \mu + \phi_p(X_{t-1} - \mu) + \phi_{p-1}(X_{t-2} - \mu) + \cdots + \phi_1(X_{t-p} - \mu),$$

where $\phi_1, ..., \phi_p$ are the solutions of the equations given in (1).

At this point we will adjust some notations: we will denote $\phi_1, ..., \phi_p$ by $\phi_{p,1}, ..., \phi_{p,p}$. The reason for this change of notation is that the coefficients for predicting X_t from $X_{t-1}, ..., X_{t-p}$ may be different the coefficients when we predict X_t from $X_{t-1}, ..., X_{t-p}, X_{t-p-1}$. So the best linear predictor of X_t using $X_{t-1}, ..., X_{t-p-1}$ is denoted by

$$\mu + \phi_{p+1,1}(X_{t-1} - \mu) + \cdot \cdots + \phi_{p+1,p}(X_{t-p} - \mu) + \phi_{p+1,p+1}(X_{t-p-1} - \mu).$$

So the partial correlation between X_t and X_{t-p-1} given $X_{t-1}, ..., X_{t-p}$ is $Corr(X_t - X_t^{(f)}, X_{t-p-1} - X_{t-p-1}^{(b)})$. It turns out that this partial correlation is equal to $\phi_{p+1,p+1}$, where $\phi_{p+1,p+1}$ is the coefficient associated with X_{t-p-1} in the best linear predictor of X_t using $X_{t-1}, ..., X_{t-p-1}$. There is an explicit formula for ϕ_{p+1} as given below

$$\phi_{p+1,p+1} = \frac{\gamma(p+1) - \phi_{p,1}\gamma(p) - \cdots - \phi_{p,p}\gamma(1)}{\gamma(0) - \phi_{p,1}\gamma(1) - \cdots - \phi_{p,p}\gamma(p)}.$$

It is important to point out that the best linear predictor (forecast) of X_t using $X_{t-1}, ..., X_{t-p-1}$ is given by

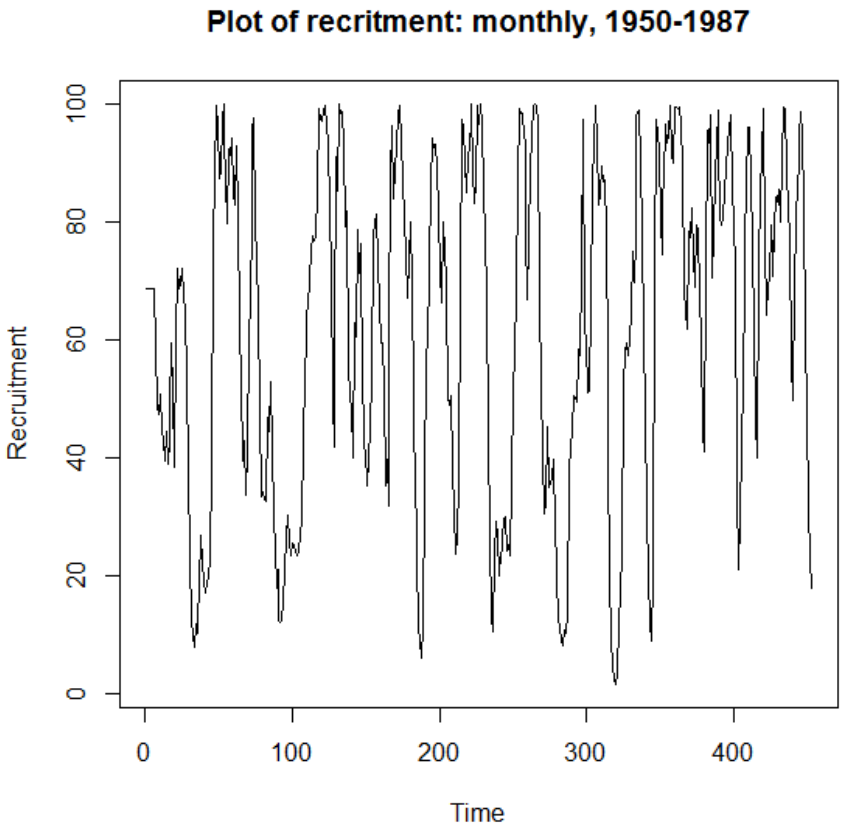
$$X_t^{(f)} = \mu + \phi_{p+1,1}(X_{t-1} - \mu) + \phi_{p+1,2}(X_{t-2} - \mu) + \cdots + \phi_{p+1,p}(X_{t-p} - \mu) + \phi_{p+1,p+1}(X_{t-p-1} - \mu),$$

where $\phi_{p+1.1}, ..., \phi_{p+1,p}, \phi_{p+1,p+1}$ are solutions of $p+1$ Yule-Walker equations

$$\begin{pmatrix} \gamma(0) & \gamma(1) & . & . & . & \gamma(p-1) & \gamma(p) \\ \gamma(1) & \gamma(0) & . & . & . & \gamma(p-2) & \gamma(p-1) \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \gamma(p-1) & \gamma(p-2) & . & . & . & \gamma(0) & \gamma(1) \\ \gamma(p) & \gamma(p-1) & . & . & . & \gamma(1) & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{p+1.1} \\ \phi_{p+1,2} \\ . \\ . \\ . \\ \phi_{p+1,p} \\ \phi_{p+1,p+1} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ . \\ . \\ . \\ \gamma(p) \\ \gamma(p+1) \end{pmatrix} \tag{2}$$

A point to note is that the the first p components (i.e., $\phi_{p+1,1}, ..., \phi_{p+1,p})$) of the solution of (2) are not the same as those obtained by solving (1). However, there is a simple relation between them. In order to write down the relation, we need a few extra notations. Let $\phi_{p,1}, ..., \phi_{p,p}$ be the solutions of (1) and let $\phi_{p+1,1}, ..., \phi_{p+1,p}, \phi_{p+1,p+1}$ be the solutions of (2). First, $\phi_{p+1,p+1}$ can be written in terms of $\phi_{p,1}, ..., \phi_{p,p}$ as follows

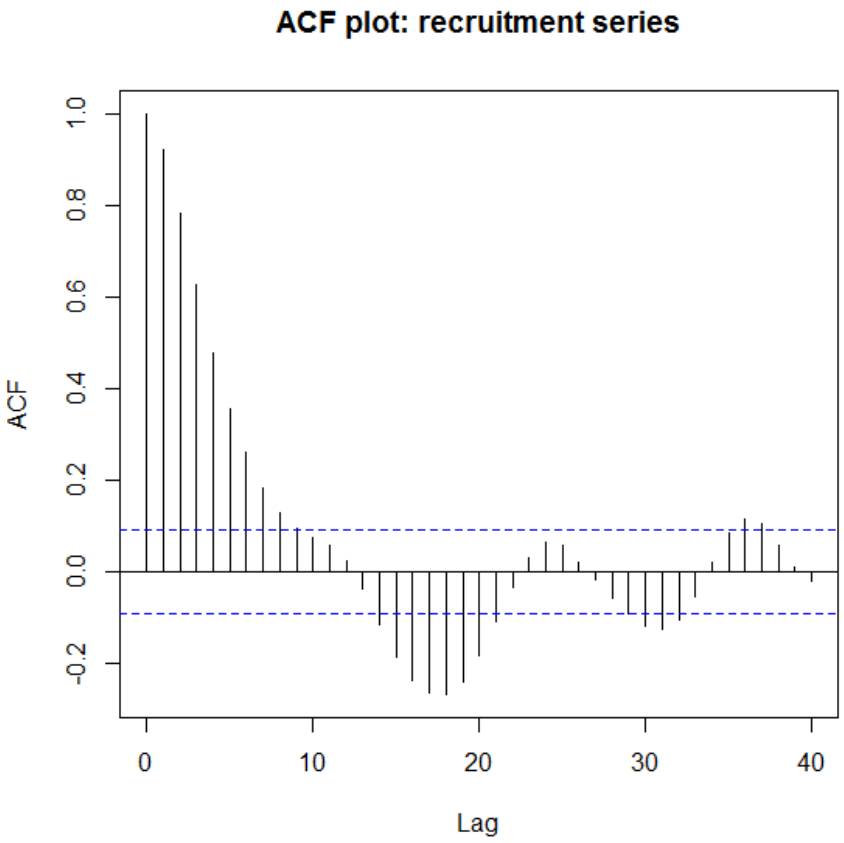
$$\begin{aligned} \phi_{p+1,p+1} &= [\gamma(p+1) - \phi_{p,1}\gamma(p) - \cdots - \phi_{p,p}\gamma(1)]/v_p, \text{ where} \\ v_p &= \gamma(0) - \phi_{p,1}\gamma(1) - \cdots - \phi_{p,p}\gamma(p). \end{aligned} \tag{3}$$

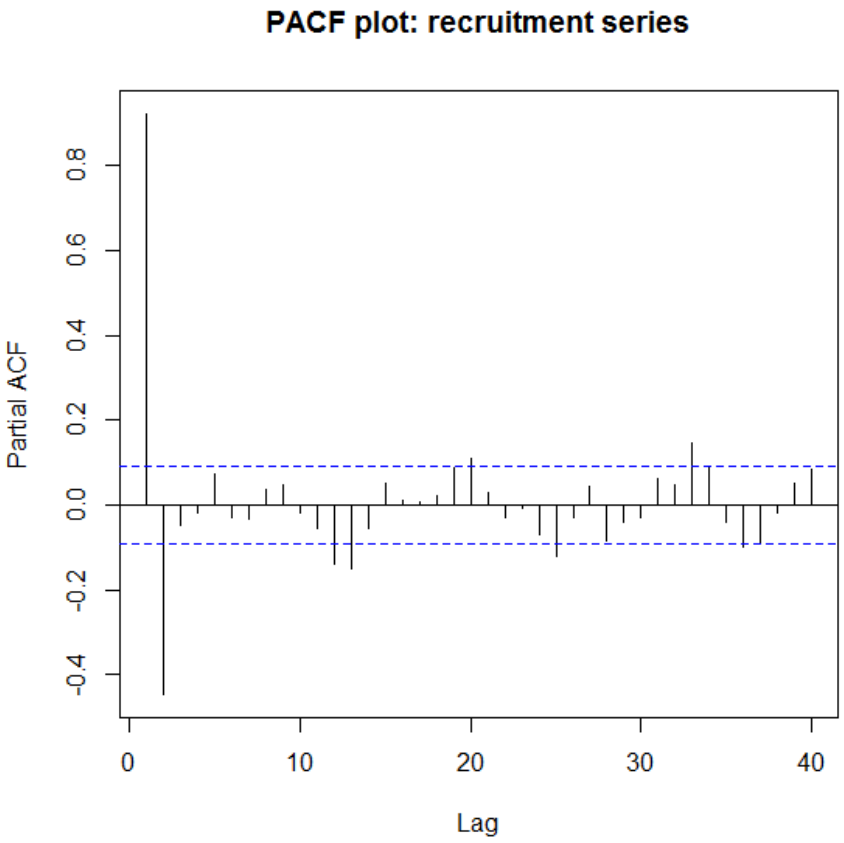


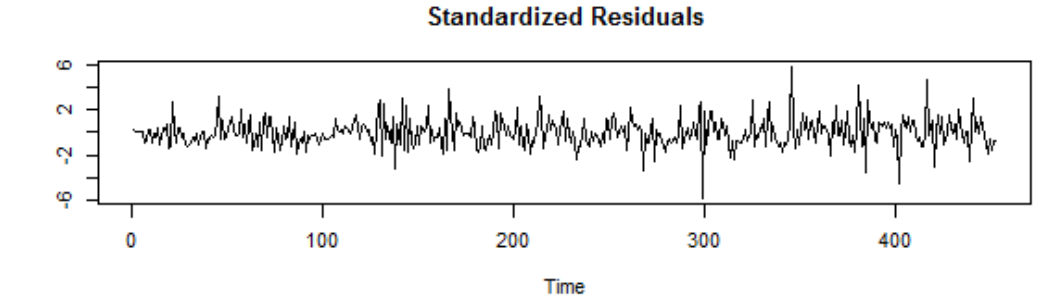
Second, then the following relation is true

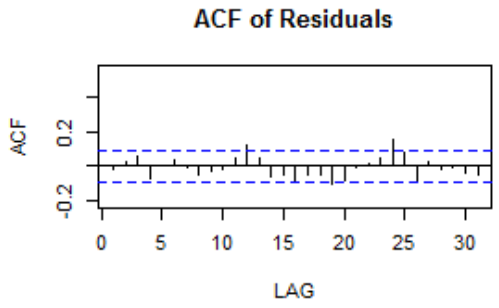
$$\begin{pmatrix} \phi_{p+1,1} \\ . \\ . \\ . \\ \phi_{p+1,p} \end{pmatrix} = \begin{pmatrix} \phi_{p,1} \\ . \\ . \\ . \\ \phi_{p,p} \end{pmatrix} - \phi_{p+1,p+1} \begin{pmatrix} \phi_{p,p} \\ . \\ . \\ . \\ \phi_{p,1} \end{pmatrix} \tag{4}$$

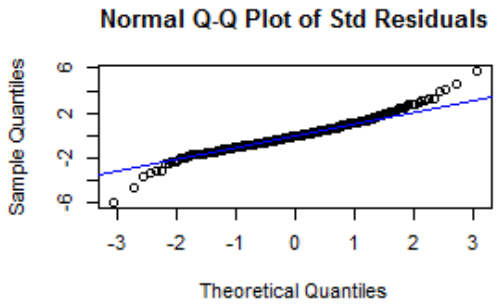
The equations (3) and (4) form the core of what is know as the Durbin-Levinson recursion formula for solving Yule-Walker equations.

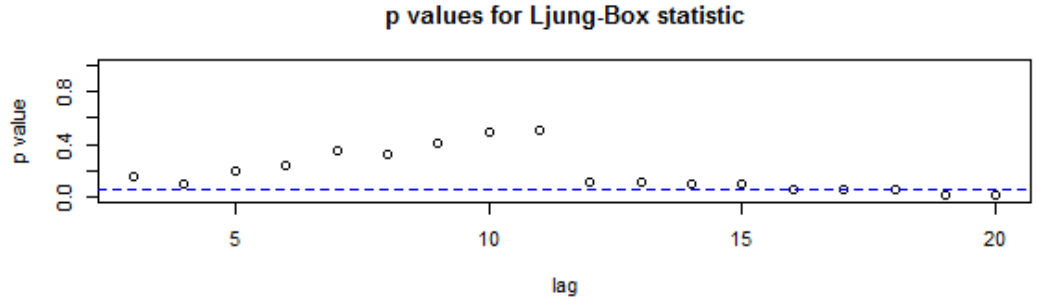

ACF plot: recruitment series
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
10
20
30
40
Lag


PACF plot: recruitment series
Partial ACF
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0
10
20
30
40
Lag


Standardized Residuals
6
2
-2
-6
0
100
200
300
400
Time


ACF of Residuals
ACF
0.2
-0.2
0
5
10
15
20
25
30
LAG


Normal Q-Q Plot of Std Residuals
Sample Quantiles
6
2
-2
-6
-3
-2
-1
0
1
2
3
Theoretical Quantiles


p values for Ljung-Box statistic
p value
0.8
0.4
0.0
5
10
15
20
lag