Week 4: SARIMA Models. Sample ACF/PACF. Model Parameter Estimation. Durbin-Levinson Algorithm.

Short Summary Of Week 3 Material:

(i) The Classical Decomposition Model: It assumes that the data is a realization of the process of the type

$$(*) \quad \boldsymbol{X_t = m_t + s_t + S_t,}$$

m_t = trend component, polynomial of order k
s_t = seasonal component, i.e. periodic with period s: $s_{t+s} = s_t$, $\sum_{j=1}^{s} s_j = 0$.
S_t = stationary process.

(ii) Differencing at lag s and an operator ∇ (nabla):
lag s difference : $W_t := \nabla_s X_t := X_t - X_{t-s} = (1 - B^s)X_t$.

lag 1 difference: $W_t := \nabla X_t \equiv \nabla^1 X_t := X_t - X_{t-1} = (1 - B)X_t$; W_t is the differenced process.

***d*th difference at lag 1:** $W_t := \nabla^d X_t = (1 - B)^d X_t$ the dth difference of X_t at lag 1.

We showed (Examples 7.1.1, 7.1.2 and 7.2.2) that

• If the trend m_t is polynomial of order d, then $W_t = \nabla^d X_t$ is stationary (differencing d times eliminates polynomial trend of order d).

• If the data is seasonal with period s, then difference at lag s, $W_t = \nabla_s X_t = (1 - B^s)X_t$, is no longer periodic (seasonal) (differencing at lag s eliminates seasonality with period s.)

• We say that a non-stationary times series $\{X_t\}$ follows **ARIMA(p,d,q) model** if $W_t = \nabla^d X_t$, produced by differencing $\{X_t\}$ d times, is a stationary ARMA(p,q).

(iii) Model for ARIMA(p,d,q) process:

$$\boldsymbol{\phi(B)(1 - B)^d X_t = \theta(B)Z_t \text{ or } \phi^*(B)X_t = \theta(B)Z_t \text{ with } \phi^*(z) = \phi(z)(1 - z)^d.}$$

Then, $\boldsymbol{W_t = \nabla^d X_t \equiv (1 - B)^d X_t}$ follows stationary ARMA(p,q) model:

$$\boldsymbol{\phi(B)W_t = \theta(B)Z_t, \text{ with } \phi(z) \neq 0, \theta(z) \neq 0 \text{ for all } |z| \leq 1.}$$

Note: X is nonstationary and the polynomial $\phi^*(z)$ has unit root $z^* = 1$ of order d.

. . . . . .

9. SARIMA models (Seasonal ARIMA) ([BD] §6.5)

SARIMA is a modification of ARIMA to account for seasonal and non-stationary behavior: if the data has seasonal behavior at lag s, the dependence on the past occurs most strongly at multiples of s. Thus, SARIMA introduces terms that identify with the seasonal lags.

To understand SARIMA:

View the time series $X_1, X_2, \ldots, X_n$ as s series: $X_j, X_{j+s}, X_{j+2s}, \ldots, X_{j+(r-1)s}$, $j = 1, 2, \ldots, s$.

For example, for monthly data, $s = 12$ and $j = 1$ may correspond to January data for r years, $j = 2$ may correspond to February data for r years, etc.:

January 1980 X_1	February 1980 X_2	March 1980 X_3	...	December 1980 X_{12}
January 1981 X_{13}	February 1981 X_{14}	March 1981 X_{15}	...	December 1981 X_{24}
...	...	...	...	...
January 2016 X_{433}	February 2016 X_{434}	March 2016 X_{435}	...	December 2016 X_{444}
$\downarrow$ X_1, X_{13}, X_{25},...	$\downarrow$ X_2, X_{14}, X_{26},...	$\downarrow$ X_3, X_{15}, X_{27},...	$\downarrow$ ...	$\downarrow$ ...

We thus have a total of s=12 series (only January or only February, etc), each has r=37 entries.

Model Assumptions:

1. Between-Year Model ARMA(P,Q). Each of these s series (that is, series taken for each month j in difference years) is generated by **the same ARMA(P, Q)** process:

$$\Phi(B^s)X_t = \Theta(B^s)U_t \text{ with } \Phi(z) = 1 - \Phi_1 z - \ldots - \Phi_P z^P, \text{ and } \Theta(z) = 1 + \Theta_1 z + \ldots + \Theta_Q z^Q.$$

2. Dependence between months ARMA(p, q). For different j's (months) U_t's are dependent and follow ARMA (p, q) process

$$\phi(B)U_t = \theta(B)Z_t,\ Z_t \sim WN(0, \sigma_Z^2) \text{ with } \phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \text{ and } \theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q.$$

If we incorporate trend, we get $\boldsymbol{\text{SARIMA}(p, d, q) \times (P, D, Q)_s}$:

$$\boldsymbol{\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t,\ \ Z_t \sim WN(0, \sigma_Z^2),\ for\ Y_t := (1-B)^d(1-B^s)^D X_t}$$

Y_t is causal if and only if $\phi(z) \neq 0, \Phi(z) \neq 0$ for $|z| \leq 1$.

We proceed with a series of examples:

Example 9.1. Pure $\text{SAR}(1)_{12}$ also written as $\text{SARIMA}(0,0,0) \times (1,0,0)_{12}$.

$s = 12, D = d = 0, P = 1, Q = 0 \Rightarrow \Phi(z) = 1 - \Phi_1 z$ **and the model is**

$$(1 - \Phi_1 B^{12})X_t = Z_t \text{ or } X_t - \Phi_1 X_{t-12} = Z_t$$

Example 9.2. Pure $\text{SMA}(1)_{12}$ also written as $\text{SARIMA}(0,0,0) \times (0,0,1)_{12}$.

$s = 12, D = d = 0, P = 0, Q = 1 \Rightarrow \Theta(z) = 1 + \Theta_1 z$ **and the model is**

$$X_t = (1 + \Theta_1 B^{12})Z_t = Z_t + \Theta_1 Z_{t-12}.$$

Note: These models correspond to data where for different j (for different months) Z_t's are uncorrelated (i.e. the s series are uncorrelated–not a very realistic assumption!)

Example 9.2-cont'd. ACVF and ACF for Pure $\text{SMA}(1)_{12}$.

For $X_t = Z_t + \Theta_1 Z_{t-12}$,

$\gamma_X(h) = E[(Z_t + \Theta_1 Z_{t-12})(Z_{t+h} + \Theta_1 Z_{t-12+h})] = \gamma_Z(h) + \Theta_1\gamma_Z(h-12) + \Theta_1\gamma_Z(h+12) + \Theta_1^2\gamma_Z(h) \neq 0$ only for $h = 0$ and $h = \pm 12$.

Conclude: The ACF and PACF for pure SMA and SAR are as follows:

P=0, Q=1, s, $X_t = Z_t + \Theta_1 Z_{t-s}$: $\rho_X(0) = 1, \rho_X(s) \neq 0$, the rest are zeros.

P=1, Q=0, s, $X_t - \Phi_1 X_{t-s} = Z_t$: $\rho_X(sk) = \Phi_1^k, k = 1, 2, \ldots$. The rest are zeros.

Example 9.3. Let $Q = 1, q = 1, P = p = 0, s = 12$. SARIMA$(0,0,1) \times (0,0,1)_{12}$ model for Y_t:

$$Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})Z_t = Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1\Theta_1 Z_{t-13}$$

This is a MA(13) model so that ACF $\rho(h) = 0$ for $h > 13$. ACF is nonzero at lags

– lag 1: $\gamma(1) = EY_tY_{t-1} = E(\ldots + \theta_1 Z_{t-1}\ldots)(Z_{t-1} + \ldots) \neq 0$,
– lag 12: $\gamma(12) = EY_tY_{t-12} = E(\ldots + \Theta_1 Z_{t-12} + \ldots)(Z_{t-12} + \ldots) \neq 0$
– lag 13: $\gamma(13) = EY_tY_{t-13} = E(\ldots + \theta_1\Theta_1 Z_{t-13})(Z_{t-13} + \ldots) \neq 0$.

However, also at lag 11 the correlation is not zero:
– $\gamma(11) = EY_tY_{t-11} = E(Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-12} + \theta_1\Theta_1 Z_{t-13})(Z_{t-11} + \theta_1 Z_{t-12} + \ldots) \neq 0$.

For a pure seasonal SMA (Q=1, q=0), only $\rho(12) \neq 0$.
When $Q = 1, q = 1$, we have nonzero ACF at lag $s + 1 = 12 + 1 = 13$ and lag $s - 1 = 11$.

Example 9.4: Let $Q = q = 0, P = p = 1, s = 12$. SARIMA$(1,0,0) \times (1,0,0)_{12}$ model for Y_t:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})Y_t = Z_t, \text{ or}$$

$$(1 - \phi_1 B - \Phi_1 B^{12} + \phi_1\Phi_1 B^{13})Y_t = Z_t \text{ or } Y_t - \phi_1 Y_{t-1} - \Phi Y_{t-12} + \phi_1\Phi_1 Y_{t-13} = Z_t$$

For example, if $\phi = .6$ and $\Phi = .5$ we have: $Y_t - .6Y_{t-1} - .5Y_{t-12} + .3Y_{t-13} = Z_t$.

This is an AR(13) model so that PACF $\alpha(h) = 0$ for $h > 13$.
PACF has distinct spikes at lags 1, 12, 13 with a bit of action coming before lag 12.

Procedure to identify SARIMA:

1. Find d, D to make $Y_t = (1 - B)^d(1 - B^s)^D X_t$ stationary. In practice usually use $d = 1, 2$ and $D = 1$.

2. Find P and Q: look at $\hat{\rho}(ks)$, $k = 1, 2, \ldots$, i.e. look at ACF and PACF at lags which are multiples of s. Identify ARMA(P, Q).

3. Find p, q: $\hat{\rho}(1), \ldots, \hat{\rho}(s - 1)$ should look as ACF of ARMA (p, q).

Note: Y_t constitutes ARMA$(p + sP, q + sQ)$ process in which some of the coefficients are zeros and the rest of the coefficients are functions of $\underline{\beta}' = (\underline{\phi}', \underline{\Phi}', \underline{\theta}', \underline{\Theta}')$.

4. Use ML Estimation for $(\underline{\beta}, \sigma_Z^2)$ and use AICC and diagnostic checking to identify the best model. (Future lectures)

10. Estimation of ACF and PACF. ([BD], §§1.4.1, 2.4, 3.2.3, 2.5.3)

In practice, time series analysis starts with a sample of consecutive observations, $x_1, \ldots, x_n$.
To choose a model, we gain information about the process from this sample.

Step 1. Plot the data and see whether the process is stationary. If not, try transformation (e.g., logarithms, powers, etc), differencing, etc. to make data stationary.

Step 2. Use stationary sample we find moment estimators.

10.1 Sample mean for stationary time series X. ([BD], §2.4.1)

Sample Mean: $\bar{X}_n = (1/n)\sum_{t=1}^n X_t$. (stationarity is required: must have $EX_t = \mu = constant$.).

Unbiased: $E(\bar{X}_n) = (1/n)\sum_{t=1}^n E(X_t) = \mu$.

Confidence Interval: For n large, distribution of the sample mean $\bar{X}_n$ is approximately normal with mean $\mu \equiv EX_t$ and variance $n^{-1}v$, where

$v = \gamma_X(0) + 2\sum_{1\le h<n}(1-\frac{h}{n})\gamma_X(h)$. For large n, one may use approximation $v = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)$.

Thus, an approximate confidence interval for μ_X is $(\bar{X}_n - 1.96v^{1/2}/\sqrt{n}, \bar{X}_n + 1.96v^{1/2}/\sqrt{n})$.

In practice, v is not known and is estimated from the data by $\hat{v} = \hat{\gamma}_X(0) + 2\sum_{1\le h<n}(1-\frac{h}{n})\hat{\gamma}_X(h)$.

10.2 Sample variance and ACF for stationary time series X.([BD], §2.4.2)

Sample Variance Estimator: $\hat{\sigma}_X^2 \equiv \hat{\gamma}_X(0) = (1/n)\sum_{t=1}^n (X_t - \bar{X}_n)^2$.

Sample ACVF at *lag h*: $\hat{\gamma}_X(h) \equiv \hat{\gamma}(h) = (1/n)\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$. (Here $\hat{\gamma} = \hat{\gamma}_X$).

Note on sample size: Use $n \ge 50,\ h \le n/4$ for these calculations (o.w. the sum has too few terms).

Sample ACF at *lag h*: $\hat{\rho}_X(h) \equiv \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$.

Characteristics of sample ACF for NON stationary process:

- For nonstationary TS usually $|\hat{\rho}(h)|$ remains large for a long time.
- For data with strong deterministic periodic component, also ACF is periodic.

10.3 Bartlett's formula: confidence intervals for sample acf. ([BD], §2.4.2– 2.4.4)

Motivation:

→ We saw that for MA(1) process $X_t = Z_t + \theta Z_{t-1}$ only ACF at lag 1 is non-zero.

→ Thus, given a sample, we estimate its ACF (find $\hat{\rho}(h)$), and if $\hat{\rho}(h), h \ge 2$ are almost zero, than we suspect that we deal with model $X_t = Z_t + \theta Z_{t-1}$.

→ However, what does it mean "almost zero"? We need to be able to write a confidence interval for $\hat{\rho}(h)$. So, we need to know distribution of $\hat{\rho}(h)$ and its variance, in particular.

The following Theorem says that the distribution of $\hat{\rho}(h)$ is Gaussian for large n and thus, "almost zero" means $|\hat{\rho}(h)| < 1.96\sqrt{Var(\hat{\rho}(h))}$.

Distribution of sample ACF. Bartlett's formula:

Assumptions:
(i) Process $X_1, \ldots, X_n$ is stationary with $Z_t \sim IID(0, \sigma_Z^2)$;
(ii) n is large.
Recommended (Box and Jenkins): $n \geq 50$, $h \leq n/4$.
Recommended (this class): $n \geq 100$ or more (b/c for $n = 50$ and $h \leq 50/4 = 12.5$ one cannot see seasonal effects.)

Main Result: Let $\underline{\rho}_h := (\rho(1), \ldots \rho(h))'$. It is estimated from a sample by the sample acf $\underline{\hat{\rho}}_h$.

The sample acf $\underline{\hat{\rho}}_h$ is approximately $\mathcal{N}(\underline{\rho}_h, n^{-1}W)$, with covariance matrix W with elements

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

Specifically, (note: math symbol $\forall$ stands for " for all" or "for every")

(10.3.1) iid noise: $\rho(k) = 0 \; \forall k \neq 0, \; \Rightarrow \; w_{ij} = 0 \, \forall i \neq j \Rightarrow Var(\hat{\rho}(h)) \approx \frac{1}{n} w_{hh} = \frac{1}{n}$, $h \geq 1$.

Conclude: for large n, for IID WN sequence, $\hat{\rho}(1), \ldots, \hat{\rho}(h)$ are approximately i.i.d. $N(0, 1/n)$.

If $|\hat{\rho}(h)| < 1.96 n^{-1/2}$ for all $h \geq 1$, assume WN (MA(0)). (95% confidence)

(10.3.2) MA (q): $X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$, $Z_t \sim IID(0, \sigma_Z^2)$. Then, $\rho(q+j) = 0$, $j \geq 1$, so that

$$Var(\hat{\rho}(h)) \approx \frac{1}{n} w_{hh} = (algebra) \, \frac{1}{n} \Big(1 + 2 \sum_{k=1}^{q} \rho^2(k) \Big), \; h > q$$

Conclude: For MA(q) sequence, and large n, $\hat{\rho}(h)$ are approximately $\mathcal{N}(\rho(h), n^{-1} w_{hh})$.

For $h > q$, $\hat{\boldsymbol{\rho}}(h)$ are approximately $\boldsymbol{\mathcal{N}}(0, \frac{1}{n}\Big(1 + 2\sum_{k=1}^{q} \rho^2(k)\Big))$.

Identification of MA(q) model, $q > 0$:
If $|\hat{\rho}(1)| > 1.96 n^{-1/2}$, then compare the rest of $\hat{\rho}(h)$ with $1.96 n^{-1/2}(1 + 2\rho(1)^2)^{1/2}$.

However, $\rho(1)^2$ is unknown. There are two possibilities to proceed:

- replace $\rho(1)$ by its estimate, i.e. check for $|\hat{\rho}(h)| < 1.96 n^{-1/2}(1 + 2\hat{\rho}(1)^2)^{1/2}$, $h \geq 2$.
  If Yes, assume MA(1) model.

- write $2\hat{\rho}(1)^2/n \sim 0$ for large n. Thus, see whether $|\hat{\rho}(h)| < 1.96 n^{-1/2}$, $h \geq 2$.
  If Yes, assume MA(1) model.

In general, if $|\hat{\rho}(h_0)| > 1.96 n^{-1/2}$ and $|\hat{\rho}(h)| < 1.96 n^{-1/2}$, $h > h_0$, assume MA(q) model with $q = h_0$.

Because we throw away nonnegative (unknown to us) terms $2\sum_{k=1}^{q} \rho^2(k)$, if a value of sample ACF $\hat{\rho}(h) \approx 1.96 n^{-1/2}$, assume that $\hat{\rho}(h)$ is within the confidence interval.

10.4 Sample PACF - Definition and distribution. ([BD], §3.2.3)

• Yule-Walker equations (4.3.1) were derived in §4.3 of Week 3. In matrix form, the equations are written as $R_h\phi_h = \rho_h$.

• in §6 of Week 3, Yule-Walker equations were used to define PACF $\alpha(h) \equiv \phi_{hh}$ as the last component of the solution vector $\underline{\phi}_h = (\phi_{h1}, \ldots, \phi_{hh})'$.

• To estimate PACF, replace unknown ACF ρ in Yule-Walker equations by its sample estimate $\hat{\rho}$:

$$\begin{aligned}
&\hat{\phi}_{h1} + \hat{\phi}_{h2}\hat{\rho}_X(1) + \ldots + \hat{\phi}_{hh}\hat{\rho}_X(h-1) = \hat{\rho}_X(1)\\
&\hat{\phi}_{h1}\hat{\rho}_X(1) + \hat{\phi}_{h2} + \ldots + \hat{\phi}_{hh}\hat{\rho}_X(h-2) = \hat{\rho}_X(2)\\
&\ldots\\
&\hat{\phi}_{h1}\hat{\rho}_X(h-1) + \hat{\phi}_{h2}\hat{\rho}_X(h-2) + \ldots + \hat{\phi}_{hh} = \hat{\rho}_X(h).
\end{aligned}$$

In matrix form: $\hat{R}_h\underline{\hat{\phi}}_h = \underline{\hat{\rho}}_h$. Here

$$\underline{\hat{\rho}}_h = (\hat{\rho}_X(1), \ldots, \hat{\rho}_X(h))', \qquad \boldsymbol{\underline{\hat{\phi}}_h = (\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh})'}, \qquad \hat{R}_h = [\hat{\rho}_X]_{i,j}.$$

• Find solution of this system of equations: $\underline{\hat{\phi}}_h = \hat{R}_h^{-1}\underline{\hat{\rho}}_h$.

Definition: Sample PACF $\hat{\alpha}_h \equiv \hat{\phi}_{hh}$ is defined as the last component of vector $\underline{\hat{\phi}}_h$.

Distribution of Sample PACF:

For **large samples**, for **AR(p)** process, the sample PACF at lags $h > p$ are approximately independent $\mathcal{N}(0, 1/n)$. (Note: $Var(\hat{\alpha}(h)) \approx 1/n$, $h > p$.)

Identification of AR(p) model:

If $|\hat{\alpha}(h)| > 1.96n^{-1/2}$ for $h = p$ and $|\hat{\alpha}(h)| \le 1.96n^{-1/2}$ for $h > p$, then suspect AR(p) model.

10.5 Calculating sample PACF: ([BD], §2.5.3)

• To calculate $\hat{\phi}_{hh}$ at a specific lag h, using the above definition, one has to solve the system of linear equations: $\hat{R}_h\hat{\phi}_h = \hat{\rho}_h$. When h runs as $h = 1, 2, \ldots, K$ with K large (e.g., **40** in all graphs in [BD]) this means the evaluation of many determinants of high dimension.

• Alternatively, we can fit AR models of increasing orders (see §6 of Week 3 Lecture Notes); then the estimate of the last coefficient in each model is the sample PACF.

Specifically, the estimates $\hat{\phi}_{hj}$ of ϕ_{hj}'s in $X_t = \phi_{h1}X_{t-1} + \ldots + \phi_{hh}X_{t-h} + Z_t$ can be updated recursively from the estimates of $\phi_{h-1,j}$ at a previous step $h-1$.
($\phi_{h-1,j}$'s correspond to $X_t = \phi_{h-1,1}X_{t-1} + \ldots + \phi_{h-1,h-1}X_{t-h+1} + Z_t$.)

Durbin-Levinson Algorithm gives the updating recursive equations:

$$\hat{\phi}_{hh} = \frac{\hat{\rho}(h) - \sum_{j=1}^{h-1}\hat{\phi}_{h-1,j}\hat{\rho}(h-j)}{1 - \sum_{j=1}^{h-1}\hat{\phi}_{h-1,j}\hat{\rho}(j)}, \; \hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh}\hat{\phi}_{h-1,h-j}, \; j = 1, \ldots, h-1.$$

([BD], §2.5.3)

These recursive equations avoid the inversion of $h \times h$ matrices.

10.5.1 Example. Assume that sample ACF $\hat{\rho}(h)$ were calculated previously.

h=1: $\hat{\phi}_{11} = \hat{\rho}(1)$

h=2: $\hat{\phi}_{22} = (\hat{\rho}(2) - \hat{\phi}_{11}\hat{\rho}(1))/(1 - \hat{\phi}_{11}\hat{\rho}(1)) = (\hat{\rho}(2) - \hat{\rho}^2(1))/(1 - \hat{\rho}^2(1))$
$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11}$

h=3: $\hat{\phi}_{33} = (\hat{\rho}(3) - \hat{\phi}_{21}\hat{\rho}(2) - \hat{\phi}_{22}\hat{\rho}(1))/(1 - \hat{\phi}_{21}\hat{\rho}(1) - \hat{\phi}_{22}\hat{\rho}(2))$
$\hat{\phi}_{31} = \hat{\phi}_{21} - \hat{\phi}_{33}\hat{\phi}_{22}$, $\hat{\phi}_{32} = \hat{\phi}_{22} - \hat{\phi}_{33}\hat{\phi}_{21}$; etc.

. . . . . .

11. Model Parameter Estimation. (Based on Appendix B and §5.1 and 5.2)

Problem: Assume that we have a stationary sample of observations $(x_1, \ldots, x_n)$ (possibly after transformation and differencing) and that we identified (using graphs of sample ACF $\hat{\rho}(h)$ and sample PACF $\hat{\phi}_{hh}$) an ARMA (p, q) model. Our next step is to estimate model parameters: $\theta_1, \ldots, \theta_q$; $\phi_1, \ldots, \phi_p$; σ_Z^2.

In this section, we discuss:
- Moment estimates known as
  - Yule-Walker estimates for pure AR and
  - Innovation estimates for pure MA and ARMA.
- Least square estimates (LSE).
- Maximum likelihood estimates (MLE).

11.1 Preliminary estimation: method of moments.

11.1.1. Yule-Walker estimation for AR(p) processes. ([BD] §5.1.1)

• For AR(p) process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t$ the ACF's $\rho(h)$'s satisfy Yule-Walker equations (see (4.3.1) and §6 of week 3 lecture notes):

$R_p \underline{\phi}_p = \underline{\rho}_p$, $\gamma(0) \equiv \sigma_X^2 = \frac{\sigma_Z^2}{1-\phi_1\rho(1)-\ldots-\phi_p\rho(p)}$.

• Replace unknown ACF's $\rho(h)$'s in Yule-Walker equations by sample ACF's $\hat{\rho}(h)$'s to get:

$\hat{R}_p \, \underline{\hat{\phi}}_p = \underline{\hat{\rho}}_p$; $\hat{\sigma}_Z^2 = \hat{\gamma}(0)\{1 - \underline{\hat{\phi}}_p' \, \underline{\hat{\rho}}_p\}$ (*)

• **Yule-Walker estimates of model parameters $\underline{\phi} = (\phi_1, \ldots, \phi_p)', \sigma_Z^2$ are solutions of equations (***

Assuming $\hat{\gamma}(0) > 0$ (to guarantee non-singularity of $\hat{R}_p$), solve for $\underline{\hat{\phi}}_p$ and $\hat{\sigma}_Z^2$:

$\underline{\hat{\phi}}_p = \hat{R}_p^{-1} \underline{\hat{\rho}}_p$, $\hat{\sigma}_Z^2 = \hat{\gamma}(0)\{1 - \underline{\hat{\rho}}_p' \hat{R}_p^{-1} \, \underline{\hat{\rho}}_p\}$.

• **Distribution of the Yule Walker estimates for large sample**: For large sample,

$\underline{\hat{\phi}}_p \approx N(\underline{\phi}, n^{-1} \sigma_Z^2 \, \Gamma_p^{-1})$. ($\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^p$ is the autocovariance matrix.)

This result allows to write (large sample) confidence intervals for ϕ, using normal distribution and estimates $\hat{\sigma}_Z^2$ and $\hat{\Gamma}_p$ for σ_Z^2 and Γ_p.

• **In practice,** to calculate $\underline{\hat{\phi}}$ recursively, without inverting $\hat{R}$ or $\hat{\Gamma}$, one uses Durbin-Levinson algorithm. (Discussed in §10.5. [BD] pp. 60-62, also 123-124.)

The algorithm iteratively calculate the vector $\underline{\hat{\phi}}_{\boldsymbol{h}} = (\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh})'$ for $\boldsymbol{h = 1, 2, \ldots}$. The last component of each vector $\hat{\phi}_{hh}$ is the sample PACF.

If $|\hat{\phi}_{hh}| < \mathbf{1.96} n^{-1/2}$ for all $h > p$, while $|\hat{\phi}_{pp}| > \mathbf{1.96} n^{-1/2}$ we choose AR(p) model
$X_t = \hat{\phi}_{p1} X_{t-1} + \ldots + \hat{\phi}_{pp} X_{t-p} + Z_t$.

• The Durbin-Levinson algorithm formulas with following modifications–
– written through autocovariances $\gamma(= \gamma(0)\rho)$ rather than autocorrelations ρ,
– using sample estimates,
– using new notation (recursive) for the estimate of the noise variance σ_Z^2:

$$\hat{v}_h = \hat{\gamma}(0)\{1 - \underline{\hat{\phi}}_h' \underline{\hat{\rho}}_h\} \equiv \hat{\gamma}(0) - \underline{\hat{\phi}}_h' \underline{\hat{\gamma}}_h \text{ (the same as } \hat{v}_h = \hat{\gamma}(0)\{1 - \underline{\hat{\rho}}_h' \hat{R}_h^{-1} \underline{\hat{\rho}}_h\})$$

are as follows:

$\hat{\phi}_{hh} = [\hat{\gamma}(h) - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j} \hat{\gamma}(h-j)] \hat{v}_{h-1}^{-1}$;
$\hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh} \hat{\phi}_{h-1,h-j}$, $j = 1, \ldots, h-1$, and
$\hat{v}_h = \hat{v}_{h-1}[1 - \hat{\phi}_{hh}^2]$.

Note: $\hat{\phi}_{11} = \hat{\rho}(1) = \hat{\gamma}(1)/\hat{\gamma}(0)$ and $\hat{v}_0 = \hat{\gamma}(0)$.

• 95% confidence intervals for ϕ_{pj}: $\hat{\phi}_{pj} \pm 1.96\, n^{-1/2}\, \hat{v}_{jj}^{1/2}$.

Here $\hat{v}_{jj}$ is the jth diagonal element of $\hat{v}_p \hat{\Gamma}_p^{-1}$ and, as above (with $h = p$), $\hat{v}_p$ is the estimate of σ_Z^2 when the order of AR was chosen to be p.

. . . . . .

Note on Durbin-Levinson algorithm and forecasting of AR(p) processes.

Assume that we observe values $X_1 = x_1, \ldots, X_n = x_n$ coming now from AR(p) process
$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t$.

To predict (forecast) future value $X_{n+1} = \phi_1 X_n + \ldots + \phi_p X_{n-p+1} + Z_{n+1}$, given the first n observations, take conditional expectation

$$\hat{X}_{n+1} = E(\phi_1 X_n + \ldots + \phi_p X_{n-p+1} + Z_{n+1} | X_1, \ldots, X_n) = \phi_1 X_n + \ldots + \phi_p X_{n-p+1} \ (**),$$

so that $X_{n+1} - \hat{X}_{n+1} = Z_{n+1}$ and the mean square error of prediction
$v_n = E(X_{n+1} - \hat{X}_{n+1})^2$ is the estimate of σ_Z^2. The Durbin-Levinson algorithm of 11.1.1. gives formulas for estimating both v_n and the coefficients ϕ in the forecast equation (**).

. . . . . .