LECTURE 12: Sample ACF and PACF

7. **Estimation of ACF and PACF.** ([BD], §§1.4.1, 2.4)

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

(i) Plot the data and see whether the process is stationary (if not, take logarithms, difference, etc.)

(ii) For stationary sample we write
Sample Mean Estimator:
$\bar{X}_n = (1/n)\sum_{t=1}^n X_t$. Unbiased: $E(\bar{X}_n) = (1/n)\sum_{t=1}^n E(X_t) = \mu$.

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

Sample autocov. function at lag h *is defined as:*
$\hat{\gamma}(h) = (1/n)\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)$.

Use $n \geq 50, \; h \leq n/4$ for these calculations (o.w. the sum has too few terms).

Sample ACF at lag h *is defined as:* $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$. Note: $\hat{\gamma}(0) = \hat{\sigma}^2$.

(ii) **Examples.** Show pictures of sample ACF for different models.

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

(iii) **Bartlett's formula**

Motivation: we saw that for TS $X_t = Z_t + \theta Z_{t-1}$ only ACF at lag 1 is non-zero. Thus, given a sample, we estimate its ACF (find $\hat{\rho}(h)$), and if $\hat{\rho}(h), h \geq 2$ 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 Thm says that the distribution of $\hat{\rho}(h)$ is Gaussian for large n and thus, "almost zero" means $|\hat{\rho}(h)| < 1.96Var(\hat{\rho}(h))$.

Bartlett's formula: Assume that n observations $X_1, \ldots, X_n$ come from a stationary TS with IID innovations Z_t's, and that n is large. Let $\underline{\hat{\rho}}_n = (\hat{\rho}(1), \ldots \hat{\rho}(n))'$ and $\underline{\rho}_n = (\rho(1), \ldots \rho(n))'$ (here $\rho(i) = \rho_X(i)$). Then $\underline{\hat{\rho}}_n$ is distributed approximately $\mathcal{N}(\underline{\rho}, n^{-1}W)$, where W is a covariance matrix with elements

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

What does it mean? CONCLUSIONS:

(a) **iid noise:** $\rho_(k) = 0$ for all k, then $w_{ij} = 0 \forall i \neq j$ and

$$Var(\hat{\rho}(h)) \approx \frac{1}{n} w_{hh} = 1/n, \; h = 1, 2, \ldots$$

(b) **MA** (q): $X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},\ Z_t \sim IID(0, \sigma_Z^2)$. Know: $\rho(q+j) = 0,\ j = 1, 2, \ldots$; that the autocorr. coefficients are zero beyond some lag q, then

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

(c) Asymptotic distribution of $\hat{\rho}(h)$ is Normal with mean $\rho(h)$ and variance given above. Recommended (Box and Jenkins) $n \geq 50,\ \ k \leq n/4$.

Examination of sample ACF:

(i) If $|\hat{\rho}(h)| < 1.96n^{-1/2}$ for all $h \geq 1$, then we assume MA(0), i.e. WN sequence.

(ii) If $|\hat{\rho}(1)| > 1.96n^{-1/2}$, then we should compare the rest of $\hat{\rho}(h)$ with $1.96n^{-1/2}(1 + 2\rho(1)^2)^{1/2}$. However, $\rho(1)^2$ is unknown. Thus, there are two possibilities to proceed:

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

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

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

Transparency.

Sample PACF – [BD], pp.94 $\hat{\alpha}_h \equiv \hat{\phi}_{hh}$ is defined as the last component of vector $\hat{\phi}_h$ which is the solution of the system of equations ($\hat{\phi}_h = \hat{R}_h^{-1}\hat{\rho}_h$).

[Solution of Yule-Walker equations where we substitute sample ACF $\hat{\rho}$ for theoretical ACF ρ].

Result: For AR(p) process, the sample PACF at lags greater than p are approximately independent Normal r.v.s with mean zero and variance $Var\hat{\phi}_{hh} \sim 1/n$. Here $n =$ sample size, large.

To decide that the value of the PACF is zero, compare it with the standard deviation. If $|\hat{\alpha}(h)| > 1.96n^{-1/2}$ for $h = p$ and $|\hat{\alpha}(h)| \leq 1.96n^{-1/2}$ for $h > p$, then the model is AR(p).