Time Series lecture 12

Partial autocorrelation & diagnostics

Sofia Olhede

EPFL

May 14, 2025

Lecture outline

1. Partial correlation
2. Partial autocorrelation
3. Diagnostics for general ARIMA models
4. Model selection

AR & Dependence

Consider an AR(1) model:

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}$$

We have $\rho_0 = 1, \rho_1 = \phi, \rho_2 = \phi^2, \ldots$

- Thus Y_t and Y_{t-2} are correlated even if we do not write Y_t in terms of Y_{t-2}.
- This follows because Y_t is specified in terms of Y_{t-1} and as Y_{t-1} is given in terms of Y_{t-2} there is correlation.
- We would therefore like to calculate the covariance of Y_t and Y_{t-2} given the effect of the intervening variable Y_{t-1}.

Partial correlation

Partial correlation

Say that we have two random variables X and Y.

- We might want to understand their correlation in order to understand something about their dependence.
- However, if they both depend on some other random variables, say $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, then this correlation may be spuriously generated by these confounders.
- One simple way to try and avoid this is partial correlation, which aims to remove this effect.
- Partial correlation is computed by fitting the best linear model for X given $\boldsymbol{Z}$, and also for Y given $\boldsymbol{Z}$, and then looking at the correlation of the residuals.

Linear prediction

In order to define partial correlation, we need to define the best linear predictor.

Definition 12.1

Consider a collection of mean-zero random variables $X_1, \ldots, X_n, Y$. The best linear predictor of Y from $X_1, \ldots, X_n$ is

$$\mathcal{P}_{X_1,\ldots,X_n}(Y) = \sum_{j=1}^{n} \beta_j X_j \tag{12.1}$$

such that the β_js minimise

$$\mathbb{E}\left[\left(Y - \mathcal{P}_{X_1,\ldots,X_n}(Y)\right)^2\right].$$

Partial correlation: formal definition

Definition 12.2 (Partial correlation)

Consider two random variables X and Y, and some confounding random variables $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, then the partial correlation of X and Y given Z is

$$\rho_{XY \bullet Z} = \operatorname{Corr}\left(X - \mathcal{P}_{\boldsymbol{Z}}(X), Y - \mathcal{P}_{\boldsymbol{Z}}(Y)\right) \tag{12.2}$$

where $\mathcal{P}_{\boldsymbol{Z}}(X)$ denotes the best linear predictor of X from $\boldsymbol{Z}$.

- If $X, Y, \boldsymbol{Z}$ are jointly Gaussian then

$$\rho_{XY \bullet Z} = \mathbb{E}\left[\operatorname{Corr}\left(X, Y \mid \boldsymbol{Z}\right)\right].$$

- In fact, in the Gaussian case,

$$\rho_{XY \bullet Z} \stackrel{a.s.}{=} \operatorname{Corr}\left(X, Y \mid \boldsymbol{Z}\right).$$

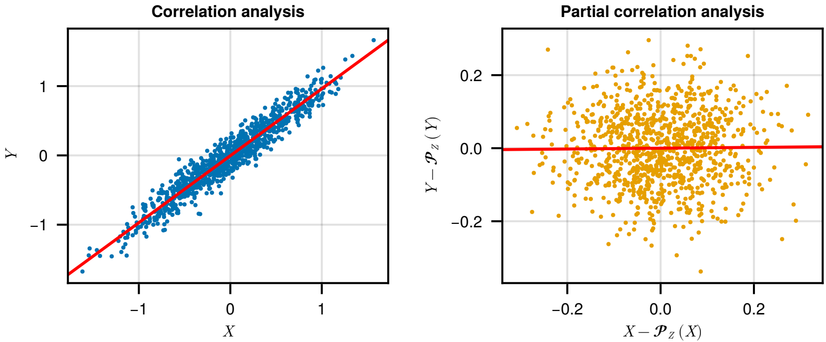


Figure: Comparison of correlation and partial correlation for some variables X and Y with some other variable Z acting as a confounder. Accounting for Z removes all correlation between X and Y, so standard correlation is misleading here!

Partial autocorrelation

Partial autocorrelation

- Partial autocorrelation is simply the partial correlation of the time series at time t with the time series at time $t+\tau$ accounting for all values in between, i.e.

$$\alpha_k = \rho_{XY\bullet \boldsymbol{Z}} \tag{12.3}$$

if we set $X = X_t,\ Y = X_{t+\tau}$ and $\boldsymbol{Z} = (X_{t+1}, \ldots, X_{t+\tau-1})^T$.

- Pictorially

$$X_t,\ \underbrace{X_{t+1}, \ldots, X_{t+\tau-1}}_{\text{confounders}},\ X_{t+\tau}.$$

Partial autocorrelation function

Definition 12.3

Consider a mean-zero stationary time series $\{X_t\}$. Denote the best linear predictors of X_t and $X_{t+\tau}$ from the intervening values

$$\hat{X}_t = \mathcal{P}_{X_{t+1},\ldots,X_{t+\tau-1}}(X_t),$$
$$\hat{X}_{t+\tau} = \mathcal{P}_{X_{t+1},\ldots,X_{t+\tau-1}}(X_{t+\tau}).$$

The partial autocorrelation function α_τ is given by

$$\alpha_\tau = \text{Corr}\left(X_{t+\tau} - \hat{X}_{t+\tau}, X_t - \hat{X}_t\right) \tag{12.4}$$

- Here we are measuring the correlation after removing the linear effects of the intervening values.

Reformulation as linear regression

Consider a mean-zero stationary time series $\{X_t\}$. Fix $t = 0$ and let $\tau \in \mathbb{Z}$, $\tau \geq 0$, the best linear predictor of $X_{t+\tau}$ from $X_t, \ldots, X_{t+\tau-1}$ takes the form

$$\mathcal{P}_{X_0,\ldots,X_{\tau-1}}(X_\tau) = \sum_{j=1}^{\tau} \alpha_{\tau,j} X_{\tau-j}. \tag{12.5}$$

- One can show that $\alpha_{\tau,\tau} = \alpha_\tau$.
- This representation is useful as we can construct a system of equations to solve for $\alpha_{\tau,\tau}$.

Relation to the ACF

For any $\tau > 0$, for $k \in \{1, \ldots, \tau\}$

$$\begin{aligned}
\gamma_k = \mathbb{E}\left[X_{\tau-k} X_\tau\right] &= \mathbb{E}\left[X_{\tau-k} \mathcal{P}_{X_0,\ldots,X_{\tau-1}}(X_\tau)\right] \\
&= \sum_{j=1}^{\tau} \alpha_{\tau,j} \mathbb{E}\left[X_{\tau-k} X_{\tau-j}\right] \\
&= \sum_{j=1}^{\tau} \alpha_{\tau,j} \gamma_{k-j}
\end{aligned}$$

and so

$$\rho_k = \sum_{j=1}^{\tau} \alpha_{\tau,j} \rho_{k-j}. \tag{12.6}$$

Thus we have equations to relate the ACF and the PACF.

ACF and PACF of ARMA models

For causal and invertible ARMA(p, q) models, the ACF and PACF have the properties

	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

Table: ACF and PACF properties of ARMA models.

Computation via Durbin-Levinson recursions

One can show that the Durbin-Levinson recursions can be used to solve for $\alpha_{\tau,\tau}$ given the ACF:

$$\alpha_{1,1} = \rho_1,$$
$$\alpha_{\tau,\tau} = \frac{\rho_\tau - \sum_{j=1}^{\tau-1} \alpha_{\tau-1,j}\rho_{\tau-j}}{1 - \sum_{j=1}^{\tau-1} \alpha_{\tau-1,j}\rho_j},$$
$$\alpha_{\tau,j} = \alpha_{\tau-1,j} - \alpha_{\tau,\tau}\alpha_{\tau-1,\tau-j}.$$

- This can be used either with the theoretical or estimated ACF.

Example time series

- A time series model has $\rho_1 = 2/5, \rho_2 = -1/20$ and $\rho_3 = -1/8$.
- Find the PACF at lags 1,2 and 3.
- We use the Durbin-Levinson recursions. These give

$$\alpha_{1,1} = \rho_1 = \frac{2}{5}$$
$$\alpha_{2,2} = \frac{\rho_2 - \alpha_{1,1}\rho_1}{1 - \alpha_{1,1}\rho_1} = -\frac{1}{4}$$
$$\alpha_{2,1} = \alpha_{1,1} - \alpha_{2,2}\alpha_{1,1} = 1/2$$
$$\alpha_{3,3} = \ldots = 0$$

We don't know about $\alpha_{k,k}$ but the latter indicates that this may be an AR(2) model.

Example time series continued

- In fact the AR(2) of

$$Y_t = \frac{1}{2} Y_{t-1} - \frac{1}{4} Y_{t-2} + \varepsilon_t$$

  has the same PACF to the process that we noted (for the first three lags).
- Note that α_{22} is equal to the coefficient of Y_{t-2} in the model.

Partial autocorrelation of an AR model

Proposition 12.4

For an AR(p), we have

$$\alpha_\tau = 0, \qquad \forall \tau > p.$$

- Furthermore, it can be shown that asymptotically the estimated partial autocorrelation at lags greater than p have mean 0 and variance $1/n$, where n is the sample size.

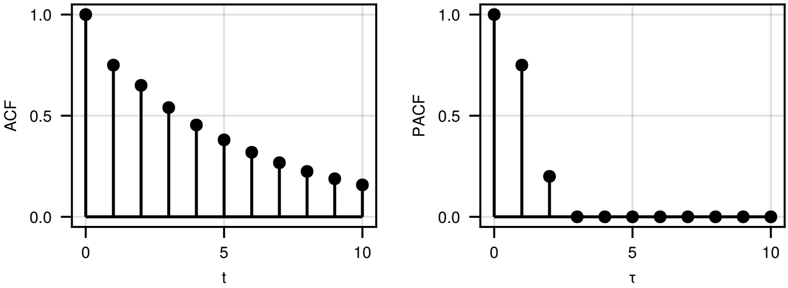


Figure: The ACF (left) and PACF (right) of model 1.

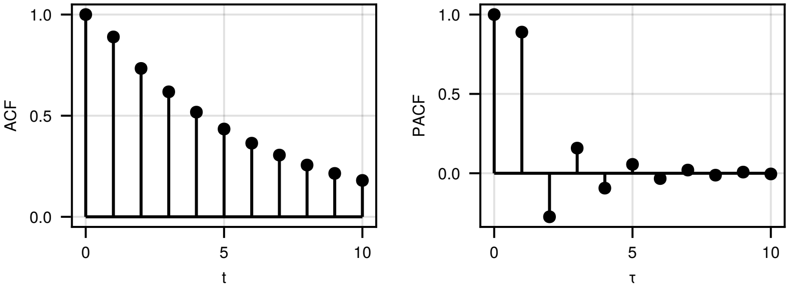


Figure: The ACF (left) and PACF (right) of model 2.

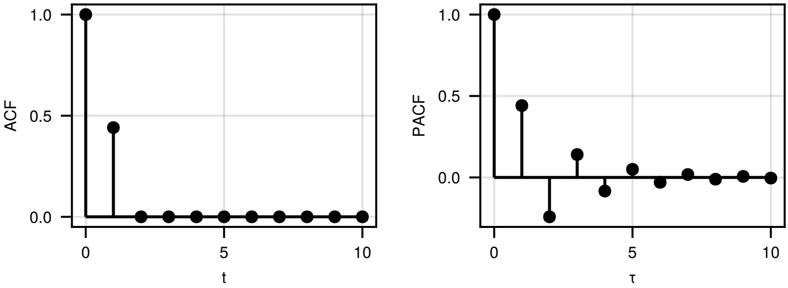


Figure: The ACF (left) and PACF (right) of model 3.

Solutions

The true models were as follows:

- model 1: AR(2)
  - $\rightarrow$ $\phi_1 = 0.6$ and $\phi_2 = 0.2$
- model 2: ARMA$(1, 2)$
  - $\rightarrow$ $\phi_1 = 0.6$, $\phi_2 = 0.2$ and $\theta_1 = -0.6$
- model 3: MA(1)
  - $\rightarrow$ $\theta_1 = -0.6$

in all cases the noise had variance 1.

Diagnostics for general ARIMA models

Residuals

Model checking is usually based on residuals. Informally they are the difference between the observed and the fitted values.

Definition 12.5 (Standardized residual)

Consider observations of a time series $\{X_t\}$. Say that for a given model we have a fitted value $\hat{X}_t$ at time t. The residuals are

$$e_t = X_t - \hat{X}_t$$

and the standardized residuals are

$$\tilde{e}_t = e_t / \sqrt{\text{Var}\,(e_t)}.$$

- Do the residuals have a constant zero mean?
- Is their variance constant wrt t (homoscedasticity)?
- Are they uncorrelated in t?
- Are they Gaussian?

Residuals for an AR(1)

Example 12.6

For an AR(1) model with parameters ϕ and σ, the fitted value at time t is ϕX_{t-1}. Therefore, the residual (if we know the true model) is given by

$$e_t = X_t - \phi X_{t-1} = \varepsilon_t.$$

This is the original noise process, meaning that the residuals have constant zero mean, constant variance and are uncorrelated. If the noise process was Gaussian, then the residuals are Gaussian.

Checking the mean

We make a plot of $\{(t, \tilde{e}_t)\}$. There should be no trends, and they should be close to zero.

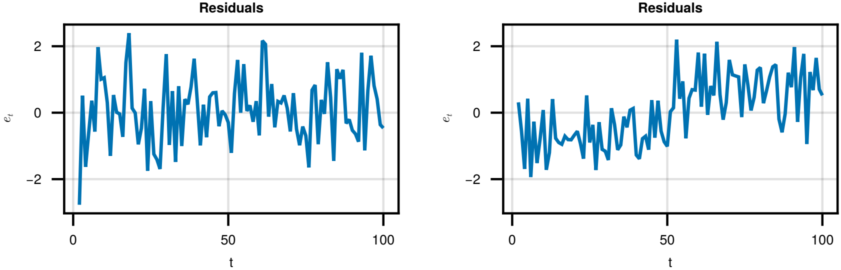


Figure: Left: centered residuals. Right: residuals with a trend.

Checking the variance

We make a plot of $\{(t, \tilde{e}_t)\}$. The variance should be constant.

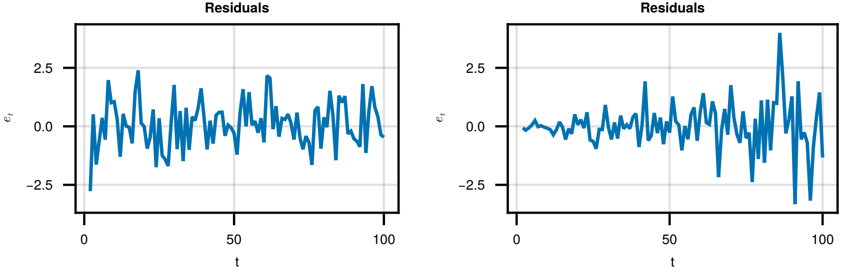
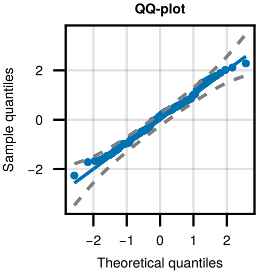


Figure: Left: constant variance. Right: non-constant variance.

Checking the Gaussianity

To check the Gaussianity of the residuals we can use a Q-Q plot. This is a plot of the quantiles of the residuals against the quantiles of a normal distribution. If the residuals are Gaussian then the points should lie on a straight line. Note that this only checks marginal Gaussianity.



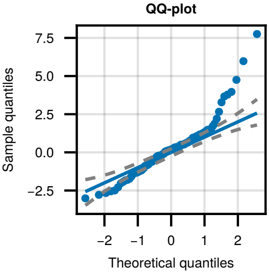


Figure: Left: Gaussian residuals. Right: student residuals.

Checking the autocorrelation

- To check the correlation we can consider the correlation between the residuals at different lags.
- Recall that the autocorrelation at lag k is given by

$$\hat{r}_\tau = \hat{\rho}_\tau^{(e)} = \frac{\sum_{t=1}^{n-\tau} \left(e_{t+\tau} - \bar{e}\right)\left(e_t - \bar{e}\right)}{\sum_{t=1}^{n} \left(e_t - \bar{e}\right)^2}.$$

- We plot the corellelogram for the residuals. Under the assumption of white noise we compare to $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$.

Testing the autocorrelation

- We can also, more realistically, try to test that a long range of correlations are zero:

$$H_0 : \quad \rho_1 = \rho_2 = \cdots = \rho_m = 0$$

for some choice of m, against the alternative that at least one autocorrelation in that range is non-zero.

Proposition 12.7 (Box-Pierce)

We introduce the Box-Pierce statistic

$$Q_m = n \sum_{\tau=1}^{m} \hat{r}_\tau^2$$

Under H_0 this is approximately χ^2_{m-p-q} for an ARMA(p, q).

Improving the Box-Pierce test

In practice this approximation is inaccurate. Thus we chose to use the modified Box-Pierce statistic, usually called the Ljung–Box statistic.

Proposition 12.8 (Ljung-Box)

$$Q_m = n(n+2)\sum_{\tau=1}^{m} \frac{\hat{r}_\tau^2}{n-\tau}$$

Under H_0 this is a χ^2_{m-p-q} for an ARMA(p, q).

- m has to be chosen by the practitioner.
- Often a series of m values are tested and displayed in a plot.

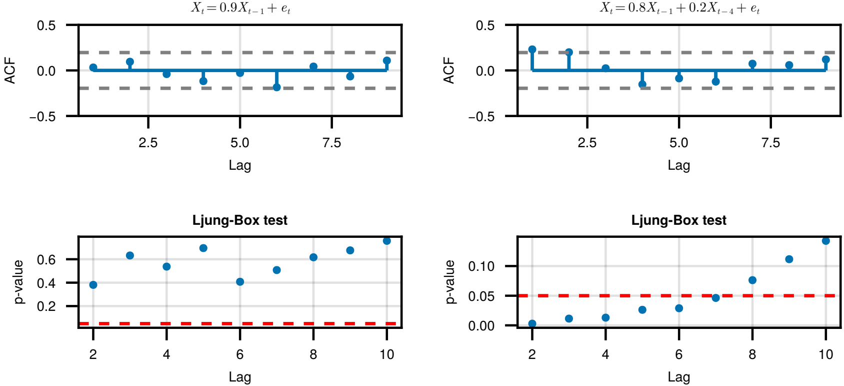


Figure: Left are the diagnostic plots for the true model; right are the diagnostic plots for an AR(2) model fitted to the same data.

Evaluating model fit: the dangers of R^2

- We might think that residual variance can help with model selection. We define the R^2 statistic to be

$$R^2 = 1 - \frac{s_e^2}{s_y^2}$$

- Some care is needed; more parameters $\Rightarrow R^2$ improves.
- Stochastic processes are not perfectly predictable. Therefore there is a limit. Consider an AR(1) model.
- The process variance is $\sigma^2 / \left(1 - \phi^2\right)$, thus

$$R^2 = 1 - \frac{\sigma^2 \left(1 - \phi^2\right)}{\sigma^2} = \phi^2$$

As $|\phi| < 1$ say with $\phi = 0.4$ we get $R^2 = 0.16$. Does not look good.

Model selection

Model Comparison

There are formal methods of comparing models.

- Most come from information theory, and correspond to information criteria.

These criteria are defined for models with k parameters

- For ARMA $k = p + q + 1$, AR and MA plus noise variance.

Information criterion are regularly used in any statistics context to compare models.

A model with smaller AIC is deemed better

$$\mathrm{AIC}(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta} \mid \mathbf{y}) + 2k$$

where L is the likelihood function.

- AIC overestimates p in the ARMA model.

Hurvich and Tsai (1989) suggested a corrected Aikake Information Criterion that works better in practice:

$$\mathrm{AICC}(\boldsymbol{\theta}) = \mathrm{AIC}(\boldsymbol{\theta}) + \frac{2k^2 - 2k}{n - k - 1}$$

- AICC and AIC become equivalent as n diverges.

The Bayesian Information Criterion (BIC) is given by

$$\mathrm{BIC}(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta} \mid \mathbf{y}) + k\log(n)$$

Box–Jenkins method

The Box-Jenkins methodology is a framework for building models:

- starts by identifying reasonable values for p, d and q,
- then estimates the parameters of the proposed ARIMA model,
- checks diagnostics to verify that the model fitting is appropriate,
- the subsequent step might be forecasting or some other inference.

Model identification

For choosing d

- Plot the data and look for non-stationarity.
- If the data looks non-stationary, plot differences of the data.
- Hopefully low orders of differencing are enough.

For choosing p and q (using the differenced data)

- Look at the acf and pacf.
- Sharp drop in the ACF at lag q suggests an MA(q) model.
- Sharp drop in the PACF at lag p suggests an AR(p) model.
- No sharp drops suggests an ARMA model.
- Very slow decay suggests non-stationary or some other issue.

Bibliography

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.