ST418 Week 9

Kaixin Liu[1]

[1]PhD Student in Statistics, LSE

Mar 19, 2026

Table of Contents

Table of Contents

Least squares estimator

For a mean-zero AR(p) model,

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \qquad t = p+1, \ldots, T.$$

Write

$$\mathbf{x}_F = \mathbf{F}\boldsymbol{\phi} + \boldsymbol{\varepsilon}_F, \qquad \mathbf{x}_F = (x_{p+1}, \ldots, x_T)^\top, \quad \boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top,$$

where $\mathbf{F}$ contains the lagged observations.

Forward least squares

$$\hat{\boldsymbol{\phi}}_F = \arg\min_{\boldsymbol{\phi}} (\mathbf{x}_F - \mathbf{F}\boldsymbol{\phi})^\top (\mathbf{x}_F - \mathbf{F}\boldsymbol{\phi}) = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{x}_F.$$

$$\hat{\sigma}_F^2 = \frac{(\mathbf{x}_F - \mathbf{F}\hat{\boldsymbol{\phi}}_F)^\top (\mathbf{x}_F - \mathbf{F}\hat{\boldsymbol{\phi}}_F)}{T - 2p}.$$

Remark: the least-squares fit need not satisfy the stationarity condition of the AR(p) model.

MLE estimator

Assume the Gaussian AR(1) model

$$x_t = \phi x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \qquad x_0 \sim N\left(0, \frac{\sigma^2}{1-\phi^2}\right).$$

Log-likelihood

$$\ell(\phi, \sigma^2) \propto -\frac{n}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}S(\phi),$$

where

$$S(\phi) = (1-\phi^2)x_0^2 + \sum_{i=1}^{n}(x_i - \phi x_{i-1})^2.$$

Exact MLE maximises $\ell(\phi, \sigma^2)$. Conditional MLE conditions on x_0 and minimises

$$S^*(\phi) = \sum_{i=1}^{n}(x_i - \phi x_{i-1})^2.$$

For large n, conditional MLE is close to exact MLE and matches least squares.

Table of Contents

AIC

Definition (Akaike's Information Criterion)

$$\text{AIC}(m) = -2\ell_{\max} + 2m,$$

where $\ell_{\max}$ is the maximised log-likelihood and m is the number of estimated parameters.

For ARMA(p, q)

Under Gaussian conditional likelihood, a common choice is

$$m = p + q + 1,$$

where the extra 1 corresponds to the innovation variance σ^2.
Fit several candidate models and choose the one with the smallest AIC.

Remark: AIC balances goodness of fit against model complexity.

Ljung-Box-Pierce

After fitting a time-series model, the residuals should behave like white noise. Let $\hat{\rho}_j$ be the sample ACF of the fitted residuals.

Portmanteau tests

$$Q^* = T(T+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{T-j}$$

is the Ljung-Box statistic, while

$$Q = T\sum_{j=1}^{k}\hat{\rho}_j^2$$

is the Box-Pierce version.

If m parameters are estimated, compare the test statistic to χ^2_{k-m}.
Remark: a large statistic or a small p-value indicates that the residuals are not white noise.

Table of Contents

Best estimator (Theorem 3.5)

Definition (MSPE)

For an ℓ-step predictor $x_t(\ell)$ of $x_{t+\ell}$,

$$\mathrm{MSPE}\big(x_t(\ell)\big) = \mathbb{E}\big[(x_t(\ell) - x_{t+\ell})^2\big].$$

Theorem (Theorem 3.5)

Among all predictors measurable with respect to the information up to time t, the conditional mean

$$x_t(\ell) = \mathbb{E}_t(x_{t+\ell})$$

minimises the mean square prediction error.

Remark: under squared-error loss, the optimal forecast is always the conditional expectation.

Prediction equations (Theorem 3.6)

Consider linear predictors of the form

$$x_t(\ell) = \alpha_0 + \sum_{k=1}^{t} \alpha_k x_k.$$

Theorem (Theorem 3.6)

$x_t(\ell)$ is the best linear predictor of $x_{t+\ell}$ if and only if

$$\mathbb{E}\big[(x_t(\ell) - x_{t+\ell})x_j\big] = 0, \qquad j = 0, 1, \dots, t,$$

where $x_0 = 1$.

After demeaning, $\alpha_0 = 0$, and for a stationary process the equations become

$$s_{t+\ell-j} = \alpha_1 s_{j-1} + \cdots + \alpha_t s_{j-t}, \qquad j = 1, \dots, t,$$

or equivalently

$$\boldsymbol{\gamma}_\ell = \boldsymbol{\Gamma}_t \boldsymbol{\alpha}, \qquad \boldsymbol{\gamma}_\ell = (s_\ell, \dots, s_{t+\ell-1})^\top, \quad \boldsymbol{\Gamma}_t = (s_{|i-j|})_{i,j=1}^{t}.$$

Table of Contents

ℓ-step prediction variance

Suppose a stationary ARMA process has MA(∞) representation

$$x_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \Psi(B)\varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2).$$

Then the best ℓ-step predictor is

$$x_t(\ell) = \mathbb{E}_t(x_{t+\ell}) = \sum_{j=0}^{\infty} \psi_{j+\ell} \varepsilon_{t-j}.$$

Hence

$$x_{t+\ell} - x_t(\ell) = \sum_{k=0}^{\ell-1} \psi_k \varepsilon_{t+\ell-k},$$

and therefore

$$\sigma^2(\ell) = \mathbb{E}\big[(x_{t+\ell} - x_t(\ell))^2\big] = \sigma^2 \sum_{k=0}^{\ell-1} \psi_k^2.$$

Remark: the first ℓ MA coefficients determine the ℓ-step forecast uncertainty.

Relating $x_t(\ell)$ to x_t

If the process is also invertible, then

$$\varepsilon_t = \Psi^{-1}(B)x_t.$$

Define

$$\Psi^{(\ell)}(z) = \sum_{j=0}^{\infty} \psi_{j+\ell} z^j, \qquad x_t(\ell) = \Psi^{(\ell)}(B)\varepsilon_t.$$

Therefore

$$x_t(\ell) = \Psi^{(\ell)}(B)\Psi^{-1}(B)x_t.$$

With ℓ fixed, the forecast sequence is itself a filtered process:

$$x_t(\ell) = G^{(\ell)}(B)x_t, \qquad G^{(\ell)}(z) = \Psi^{(\ell)}(z)\Psi^{-1}(z).$$

Remark: in practice, plug in parameter estimates and truncate the infinite filter at a sufficiently large lag.