

ST326 Week 5

Kaixin Liu[1]

[1]PhD Student in Statistics, LSE

Oct 30, 2026

Table of Contents

1 Model selection

Table of Contents

1 Model selection

MA(q): definition and ACF cutoff

Definition (MA(q))

With $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$,

$$X_t = \mu + \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}.$$

Population ACF

The ACF $\rho(k) = \mathrm{Cov}(X_t, X_{t-k})/\mathrm{Var}(X_t)$ satisfies

$$\boxed{\rho(k) = 0 \quad \text{for } |k| > q}.$$

Observations more than q periods apart share no shocks. The population ACF cuts off at q.

MA(q): sampling variation beyond the cutoff

Assume IID innovations with finite fourth moment. For fixed $k > q$,

Asymptotic distribution

$$\sqrt{T}\,\hat{\rho}(k) \xrightarrow{d} N\left(0, 1 + 2\sum_{j=1}^{q} \rho(j)^2\right).$$

95% bounds

When q is unknown, use lag-dependent bounds:

$$\pm\frac{1.96}{\sqrt{T}}\left(1 + 2\sum_{j=1}^{k-1} \hat{\rho}(j)^2\right)^{1/2}.$$

Beyond the cutoff, most estimates lie within these pointwise bounds.

AR(p) and the partial autocorrelation

Definition (AR(p))

For a stationary series, with $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$,

$$X_t = c + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \varepsilon_t.$$

Definition (PACF)

For $k \geq 1$, let $(b_{k0}, \ldots, b_{kk})$ minimise

$$\mathbb{E}\left(X_{k+1} - \beta_0 - \sum_{j=1}^{k} \beta_j X_{k+1-j} \right)^2.$$

Define $\pi(k) = b_{kk}$, the coefficient of the oldest regressor X_1.

$\pi(k)$ measures the extra linear contribution of lag k after accounting for the intervening lags.

AR(p): population and sample PACF cutoff

Population PACF

For a stationary AR(p) driven by innovations,

$$\boxed{\pi(k) = 0 \quad \text{for } k > p}.$$

Beyond p, an older lag adds nothing to the best linear predictor.

Sample PACF

Estimate $\hat{\pi}(k) = \hat{b}_{kk}$ by least squares. With IID innovations and finite fourth moment, for fixed $k > p$,

$$\sqrt{T}\,\hat{\pi}(k) \xrightarrow{d} N(0,1).$$

Approximate pointwise 95% bounds are $\pm 1.96/\sqrt{T}$.

Look for a sample PACF cutoff at p.

ARMA(p, q): remove AR dependence

Definition (ARMA(p, q))

$$X_t = c + \sum_{j=1}^{p} b_j X_{t-j} + \varepsilon_t + \sum_{j=1}^{q} a_j \varepsilon_{t-j},$$

Assume stationarity and $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$, with no cancelling factors.

AR filtering

For the centred series $Y_t = X_t - \mathbb{E}(X_t)$,

$$Z_t = Y_t - \sum_{j=1}^{p} b_j Y_{t-j} \sim \mathrm{MA}(q).$$

Removing the AR component leaves an MA process. EACF searches over AR orders to reveal its MA cutoff.

Extended autocorrelation function (EACF)

Definition (Sample EACF)

For AR row i and MA column j, use iteratively estimated coefficients:

$$\hat{Z}_t^{(i,j)} = Y_t - \sum_{\ell=1}^{i} \hat{b}_\ell^{(i,j)} Y_{t-\ell}, \qquad \hat{\rho}(i,j) = \hat{\rho}_{\hat{Z}^{(i,j)}}(j+1).$$

$\hat{\rho}(i,j)$ is the lag-$(j+1)$ sample ACF of the filtered series.

- AR estimates are adjusted for MA dependence.
- At row p, zeros begin at column q. In row $i > p$, the start shifts to $q + (i - p)$, producing a triangle.

Column $j = q$ checks lag $q + 1$, where an MA(q) ACF is zero.

EACF: zero triangle and sampling variation

AR \ MA	$q-1$	q	$q+1$	$q+2$	$\cdots$
p	x	o	o	o	$\cdots$
$p+1$	x	x	o	o	$\cdots$
$p+2$	x	x	x	o	$\cdots$

The triangle's upper-left vertex gives (p, q).

Sampling approximation

At a zero-pattern entry, for large T,

$$\hat{\rho}(i,j) \dot{\sim} N\left(0, \frac{1}{T-i-j}\right), \qquad \text{o if } |\hat{\rho}(i,j)| \leq \frac{1.96}{\sqrt{T-i-j}}.$$

Otherwise mark x. These are approximate pointwise bounds.

Normal approximation: Cryer & Chan (2008), §6.2; threshold: Chapter 2, slide 29.

AR, MA and ARMA: comparison

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cutoff at q	Tails off
PACF	Cutoff at p	Tails off	Tails off
EACF triangle vertex	$(p, 0)$	$(0, q)$	(p, q)

"Tails off" means no finite cutoff in general.

Information criteria

For a fitted stationary ARMA(p, q), let $\hat{\sigma}^2$ estimate the innovation variance and let T be the sample size.

AIC, BIC and AICC

$$\begin{aligned}
\mathrm{AIC}(p,q) &= T\log(\hat{\sigma}^2) + 2(p+q+1), \\
\mathrm{BIC}(p,q) &= T\log(\hat{\sigma}^2) + (p+q+1)\log T, \\
\mathrm{AICC}(p,q) &= T\log(\hat{\sigma}^2) + \frac{2(p+q+1)T}{T-p-q-2}.
\end{aligned}$$

- Minimise the chosen criterion: balance fit against complexity.
- BIC penalises complexity more heavily when $\log T > 2$; AICC corrects AIC for small samples ($T > p+q+2$).
- Compare models fitted to the same data with the same differencing order and estimation method.