

Econometrics

Practical Session 17

Lag Length Selection and Model Comparison

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics



Theoretical Wrap-up

How Many Lags? The Model Selection Problem

- **More lags:**
 - **Reduce bias** → we capture more autocorrelation structure
 - **Increase variance** → more parameters to estimate from finite data

How Many Lags? The Model Selection Problem

- **More lags:**
 - **Reduce bias** → we capture more autocorrelation structure
 - **Increase variance** → more parameters to estimate from finite data
- **Sequential testing approach:** start with a large p and remove lags that are individually insignificant → **Problem:** overfitting models

How Many Lags? The Model Selection Problem

- **More lags:**
 - **Reduce bias** → we capture more autocorrelation structure
 - **Increase variance** → more parameters to estimate from finite data
- **Sequential testing approach:** start with a large p and remove lags that are individually insignificant → **Problem:** overfitting models
- **Information Criteria Approach:** choose p to minimize a penalized-fit measure → the improvement in fit (\downarrow residual variance) must be worth the cost of extra parameters

Information Criteria: AIC and BIC

Let $\hat{\sigma}^2 = SSR / T$ and $p + 1$ be the AR coefficients plus the intercept. Then:

Information Criteria: AIC and BIC

Let $\hat{\sigma}^2 = SSR / T$ and $p + 1$ be the AR coefficients plus the intercept. Then:

Akaike Information Criterion

$$AIC = \ln \hat{\sigma}^2 + 2 \frac{p + 1}{T}$$

- Penalizes each extra parameter by $2/T$
- Tends to overfit models in large samples
- **In R:** `AIC(model)`

Let $\hat{\sigma}^2 = SSR / T$ and $p + 1$ be the AR coefficients plus the intercept. Then:

Akaike Information Criterion

$$AIC = \ln \hat{\sigma}^2 + 2 \frac{p + 1}{T}$$

- Penalizes each extra parameter by $2/T$
- Tends to overfit models in large samples
- **In R:** AIC(model)

Bayesian Information Criterion

$$BIC = \ln \hat{\sigma}^2 + \frac{p + 1}{T} \ln T$$

- Penalizes by $\ln(T)/T \rightarrow$ **heavier** for $T > e^2 \approx 7.4$
- Is consistent: it selects the true model as $T \rightarrow +\infty$
- **In R:** BIC(model)

MA Models (Brief Introduction)

- **Moving Average MA(q)**: regress y_t on **past shocks**

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

MA Models (Brief Introduction)

- **Moving Average MA(q)**: regress y_t on **past shocks**

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

- Always stationary
- ACF cuts off at lag q → shocks have memory lasting only q periods
- **In R**: `arima(y, order = c(0, 0, q))`

ARMA Models (Brief Introduction)

- **ARMA(p, q)**: combines both components for **parsimony**

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

ARMA Models (Brief Introduction)

- **ARMA(p, q)**: combines both components for **parsimony**

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

- An $\text{AR}(\infty)$ can often be well-approximated by a compact $\text{ARMA}(p, q)$
- Both AIC and BIC can be used to choose p and q jointly
- **In R**: `auto.arima()` in the `forecast` package selects all orders automatically

ACF and PACF as Diagnostic Tools

- **Autocorrelation Function (ACF):** $\hat{\rho}_j \rightarrow$ correlation between y_t and y_{t-j}
 - Decays geometrically for $AR(p)$
 - Cuts off sharply after lag q for $MA(q)$

ACF and PACF as Diagnostic Tools

- **Autocorrelation Function (ACF):** $\hat{\rho}_j \rightarrow$ correlation between y_t and y_{t-j}
 - Decays geometrically for $AR(p)$
 - Cuts off sharply after lag q for $MA(q)$
- **Partial Autocorrelation Function (PACF):** $\hat{\phi}_{kj}$
 - Correlation between y_t and y_{t-j} in a model with k lags, **after removing the effect of lags $1, 2, \dots, j - 1$**
 - Indicates the **direct predictive contribution of each lag** \rightarrow cuts off after lag p for $AR(p)$
 - Decays geometrically for $MA(q)$

ACF and PACF as Diagnostic Tools

Box-Jenkins Identification Table

Pattern	ACF	PACF
$AR(p)$	Geometric decay	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Geometric decay
$ARMA(p,q)$	Geometric decay	Geometric decay
White noise	All near zero	All near zero

ACF and PACF as Diagnostic Tools

Practical use of the **Box-Jenkins Identification Table**:

- Inspect **ACF/PACF of the residuals** → significant spikes signal under-fitting
- The **PACF spike at lag j** suggests adding the j -th lag to the AR model

Exercises

Exercise 1 | IP Growth and Lag Selection

(Exercise 15.2 from S&W)

The Index of Industrial Production (IP_t) measures monthly US industrial output.
Define:

$$Y_t = 1200 \times \ln\left(\frac{IP_t}{IP_{t-1}}\right)$$

All regressions are estimated over 1986:M1–2017:M12 ($T = 384$).

Exercise 1 | IP Growth and Lag Selection

a) A forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

Exercise 1 | IP Growth and Lag Selection

a) A forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

Exercise 1 | IP Growth and Lag Selection

a) A forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

KEY TAKEAWAYS

- $\ln(IP_t/IP_{t-1}) \approx (IP_t - IP_{t-1})/IP_{t-1}$: the monthly fractional change
- Multiplying by 100 gives the monthly percentage change
- Multiplying by 1200 = 100×12 **annualizes** it: if IP grew at the same rate every month, the yearly growth would be approximately $Y_t\%$
- **The statement is correct**

Exercise 1 | IP Growth and Lag Selection

b) The following AR(4) is estimated:

$$\hat{Y}_t = \underset{(0.488)}{0.749} + \underset{(0.088)}{0.071}Y_{t-1} + \underset{(0.053)}{0.170}Y_{t-2} + \underset{(0.078)}{0.216}Y_{t-3} + \underset{(0.064)}{0.167}Y_{t-4}$$

The monthly IP values for late 2017 are:

Date	2017:M7	2017:M8	2017:M9	2017:M10	2017:M11	2017:M12
IP	105.01	104.56	104.82	106.58	106.86	107.30

These give the following annualised growth rates (Y_t):

Month	Calculation	Y_t
Sep 2017	$1200 \times \ln(104.82/104.56)$	2.98

Exercise 1 | IP Growth and Lag Selection

Month	Calculation	Y_t
Oct 2017	$1200 \times \ln(106.58/104.82)$	19.98
Nov 2017	$1200 \times \ln(106.86/106.58)$	3.15
Dec 2017	$1200 \times \ln(107.30/106.86)$	4.93

The resulting one-step-ahead forecast is $\hat{Y}_{\text{Jan 2018}} \approx 6.4\%$. Which lag contributes most to this forecast? Do the signs of all four contributions make economic sense?

Exercise 1 | IP Growth and Lag Selection

Exercise 1 | IP Growth and Lag Selection

KEY TAKEAWAYS

- Contributions of each lag to the forecast:

Lag	Month	Coeff $\times Y_t$	Contribution
Y_{t-1}	Dec 2017	0.071×4.93	0.35
Y_{t-2}	Nov 2017	0.170×3.15	0.54
Y_{t-3}	Oct 2017	0.216×19.98	4.31
Y_{t-4}	Sep 2017	0.167×2.98	0.50

- The October lag dominates:** it accounts for 4.31 of the 6.4% forecast
- All coefficients are positive and all Y_t values are positive \rightarrow all contributions are positive and in the expected direction
- But the forecast is unusual: it is almost entirely driven by a hurricane-induced rebound, not genuine momentum

Exercise 1 | IP Growth and Lag Selection

***b')** Why is October 2017 ($Y_t \approx 20\%$) such an outlier? How does it affect the forecast?*

Exercise 1 | IP Growth and Lag Selection

***b')** Why is October 2017 ($Y_t \approx 20\%$) such an outlier? How does it affect the forecast?*

Exercise 1 | IP Growth and Lag Selection

***b')** Why is October 2017 ($Y_t \approx 20\%$) such an outlier? How does it affect the forecast?*

KEY TAKEAWAYS

- **Cause:** Hurricanes Harvey and Irma disrupted production in Texas, Louisiana, and Florida
- **Effect on forecast:** it contributes $0.216 \times 19.98 \approx 4.3$ pp to the January forecast → without it, the forecast would be 2.9% instead of 6.4%
- **Lesson:** AR models cannot distinguish a one-off rebound from persistent momentum — they mechanically extrapolate any past spike forward

Exercise 1 | IP Growth and Lag Selection

c) She adds Y_{t-12} to the AR(4). The estimated coefficient is -0.061 with standard error 0.043 , giving a t -statistic of -1.42 . Is this coefficient statistically significant at the 5% level? What does this tell us about annual seasonality in IP growth?

Exercise 1 | IP Growth and Lag Selection

c) She adds Y_{t-12} to the AR(4). The estimated coefficient is -0.061 with standard error 0.043 , giving a t -statistic of -1.42 . Is this coefficient statistically significant at the 5% level? What does this tell us about annual seasonality in IP growth?

Exercise 1 | IP Growth and Lag Selection

c) She adds Y_{t-12} to the AR(4). The estimated coefficient is -0.061 with standard error 0.043 , giving a t -statistic of -1.42 . Is this coefficient statistically significant at the 5% level? What does this tell us about annual seasonality in IP growth?

KEY TAKEAWAYS

- $|t| = 1.42 < 1.96 \rightarrow$ fail to reject $H_0 : \beta_{12} = 0$ at 5%
- The 12th lag is **not statistically significant**
- **Economic interpretation:** there is no strong evidence that IP growth in a given month systematically predicts growth 12 months later — annual seasonality is not a major driver of this series

Exercise 1 | IP Growth and Lag Selection

d) To select the lag order, she estimates $AR(p)$ for $p = 0, 1, \dots, 6$ and records the SSR. Using $\hat{\sigma}^2 = SSR / T$ with $T = 384$, the AIC and BIC are computed for each model:

Order p	0	1	2	3	4	5	6
SSR	21 045	20 043	18 870	17 838	17 344	17 337	17 306
BIC	4.04	4.00	3.96	3.92	3.90	3.92	3.93
AIC	4.02	3.98	3.93	3.88	3.85	3.86	3.86

Both criteria select AR(4). What does it tell you that AIC and BIC agree here? How would your conclusion change if AIC had selected AR(5) while BIC chose AR(4)?

Exercise 1 | IP Growth and Lag Selection

Exercise 1 | IP Growth and Lag Selection

KEY TAKEAWAYS

- **Both select AR(4):** the SSR drops sharply up to $p = 4$ and then barely changes → the 5th and 6th lags add almost no explanatory power
- **Agreement strengthens confidence:** AIC tends to over-select (penalizes less), so when BIC agrees with AIC's choice, the case for that order is robust
- **If they disagreed (AIC → AR(5), BIC → AR(4)):**
 - **Prefer BIC in large samples** → it is consistent and penalizes complexity more heavily
 - The **difference in fit between AR(4) and AR(5) is negligible** (SSR drops by only 7 out of 17 344), so the simpler model is preferred

Exercise 1 | IP Growth and Lag Selection

e) What does the AR(4) structure tell us economically about how shocks to industrial production propagate over time?

Exercise 1 | IP Growth and Lag Selection

e) What does the AR(4) structure tell us economically about how shocks to industrial production propagate over time?

Exercise 1 | IP Growth and Lag Selection

e) What does the AR(4) structure tell us economically about how shocks to industrial production propagate over time?

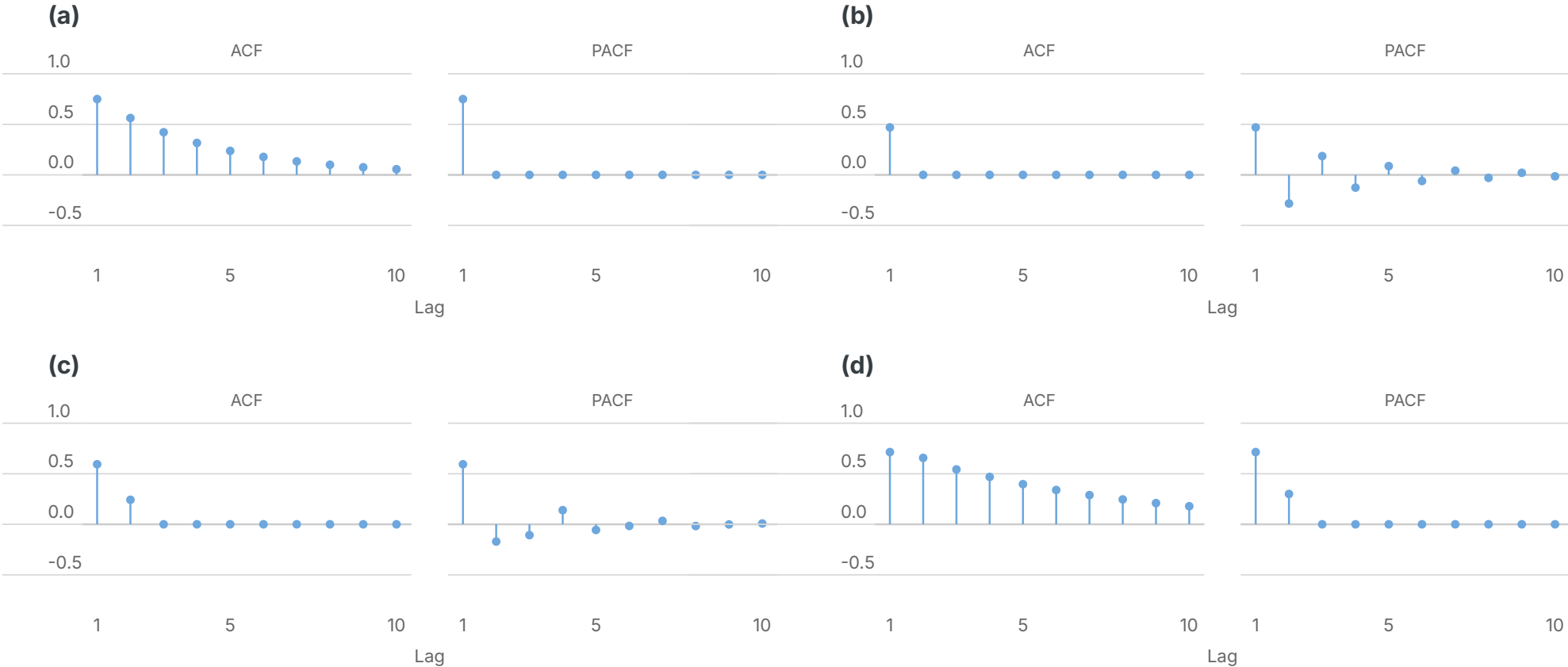
KEY TAKEAWAYS

- **Industrial production is autocorrelated over roughly 4 months** → a shock today has predictable effects for the next quarter
- This reflects real economic mechanisms: production plans, inventory adjustment, and supply chain lags take time to work through
- **Beyond 4 months**, the process is close to white noise → the shock has been absorbed

Exercise 2 | ACF/PACF Model Identification

For each process (a)–(d), identify the most likely model (AR, MA, ARMA, or white noise) and justify.

Exercise 2 | ACF/PACF Model Identification



Exercise 2 | ACF/PACF Model Identification

Exercise 2 | ACF/PACF Model Identification

KEY TAKEAWAYS

- **(a) AR(1):** PACF cuts off sharply after lag 1 → only the first lag has direct predictive power; ACF decays geometrically
- **(b) MA(1):** ACF cuts off sharply after lag 1 → shocks persist for only 1 period; PACF decays geometrically (alternating sign)
- **(c) MA(2):** ACF cuts off after lag 2 → shocks persist for 2 periods; PACF decays geometrically
- **(d) AR(2):** PACF cuts off after lag 2 → two lags have direct predictive power; ACF decays geometrically

Exercise 3 | Residual ACF Diagnostics

After fitting AR models of different orders to the IP growth series, an analyst inspects the ACF of the residuals up to lag 6. The table below marks each lag as significant (✓) or not at the 5% level:

Model	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6
AR(1)	—	✓	✓	✓	—	—
AR(2)	—	—	✓	✓	—	—
AR(4)	—	—	—	—	—	—

Exercise 3 | Residual ACF Diagnostics

a) The AR(1) residuals show significant spikes at lags 2, 3, and 4. What does this tell you about the model? How does the Box-Jenkins table guide you?

Exercise 3 | Residual ACF Diagnostics

a) The AR(1) residuals show significant spikes at lags 2, 3, and 4. What does this tell you about the model? How does the Box-Jenkins table guide you?

Exercise 3 | Residual ACF Diagnostics

a) The AR(1) residuals show significant spikes at lags 2, 3, and 4. What does this tell you about the model? How does the Box-Jenkins table guide you?

KEY TAKEAWAYS

- Significant residual autocorrelation means the model has not captured all the serial dependence in Y_t → **the AR(1) is under-fitted**
- The PACF of the residuals would show spikes at lags 2, 3, 4 → the Box-Jenkins table points to an **AR process** with direct predictive power at those lags
- **Action: add lags up to at least 4** → consistent with the BIC result from Exercise 1

Exercise 3 | Residual ACF Diagnostics

b) The AR(2) residuals still show spikes at lags 3 and 4. A student says: “*At least it is better than AR(1).*” Is that a sufficient reason to stop at AR(2)?

Exercise 3 | Residual ACF Diagnostics

b) The AR(2) residuals still show spikes at lags 3 and 4. A student says: “*At least it is better than AR(1).*” Is that a sufficient reason to stop at AR(2)?

Exercise 3 | Residual ACF Diagnostics

b) The AR(2) residuals still show spikes at lags 3 and 4. A student says: “*At least it is better than AR(1).*” Is that a sufficient reason to stop at AR(2)?

KEY TAKEAWAYS

- **Better than AR(1) is not the same as adequate:** significant residual autocorrelation at lags 3–4 means the model is still mis-specified
- A model with serially correlated residuals has **invalid standard errors and unreliable forecasts**
- **The right stopping rule:** keep adding lags until **the residual ACF shows no significant spikes**, or **until BIC stops improving**
- Here **both criteria point to AR(4)**: residuals are clean and BIC is minimized

Exercise 3 | Residual ACF Diagnostics

c) The AR(4) residuals show no significant spikes up to lag 6. A classmate concludes: *"This proves AR(4) is the correct model."* What is wrong with this statement? What else should you check?

Exercise 3 | Residual ACF Diagnostics

c) The AR(4) residuals show no significant spikes up to lag 6. A classmate concludes: *"This proves AR(4) is the correct model."* What is wrong with this statement? What else should you check?

Exercise 3 | Residual ACF Diagnostics

c) The AR(4) residuals show no significant spikes up to lag 6. A classmate concludes: *“This proves AR(4) is the correct model.”* What is wrong with this statement? What else should you check?

KEY TAKEAWAYS

- **Absence of residual autocorrelation is necessary but not sufficient** → a white-noise residual ACF means we are done modelling autocorrelation, it does not mean we are done modelling
- **What else to check:**
 - **Normality:** heavy tails or outliers (like the hurricane spike) can inflate forecast intervals
 - **Heteroskedasticity:** if residual variance shifts over time (e.g. higher volatility in recessions), the model uncertainty is mis-stated
 - **Structural stability:** if the AR coefficients changed at some point in the sample, the “clean” residuals may mask a regime shift

Discussion

1. You have an AR(2) with $BIC = 4.12$ and an ARMA(1,1) with $BIC = 4.08$. Both have two parameters. What does the lower BIC of the ARMA tell you? Would you automatically switch to ARMA?

Discussion

1. You have an AR(2) with BIC = 4.12 and an ARMA(1,1) with BIC = 4.08. Both have two parameters. What does the lower BIC of the ARMA tell you? Would you automatically switch to ARMA?

1. You have an AR(2) with BIC = 4.12 and an ARMA(1,1) with BIC = 4.08. Both have two parameters. What does the lower BIC of the ARMA tell you? Would you automatically switch to ARMA?

KEY TAKEAWAYS

- Lower BIC means the ARMA(1,1) fits better per parameter than the AR(2): it achieves lower residual variance with the same complexity
- ARMA(1,1) can sometimes represent a much higher-order AR model compactly → parsimony
- **Caveat:** ARMA models are harder to estimate (MLE, not OLS), harder to interpret, and can have identification problems
- **Practical rule:** if the BIC difference is small (< 2), stick with the AR for simplicity; switch to ARMA only if the difference is meaningful and the model estimates cleanly

2. Think about the research project. You are choosing the lag order for an AR model of Portuguese electricity consumption using daily data. Would you use a grid of $p = 1, \dots, 7$ days and pick by BIC? What pitfalls should you watch for?

2. Think about the research project. You are choosing the lag order for an AR model of Portuguese electricity consumption using daily data. Would you use a grid of $p = 1, \dots, 7$ days and pick by BIC? What pitfalls should you watch for?

2. Think about the research project. You are choosing the lag order for an AR model of Portuguese electricity consumption using daily data. Would you use a grid of $p = 1, \dots, 7$ days and pick by BIC? What pitfalls should you watch for?

KEY TAKEAWAYS

- Daily electricity consumption has a strong **weekly** cycle → lag 7 is a natural predictor
- Simply running AR(1) through AR(7) and picking by BIC may drop lag 7 if the fit improvement is small, missing the seasonal pattern entirely
- **Better approach:** always include lag 7 regardless of BIC, then use BIC to choose among shorter lags
- **Pitfall:** if the series also shows annual seasonality, lag 365 matters too, but estimating it requires at least a couple of years of data