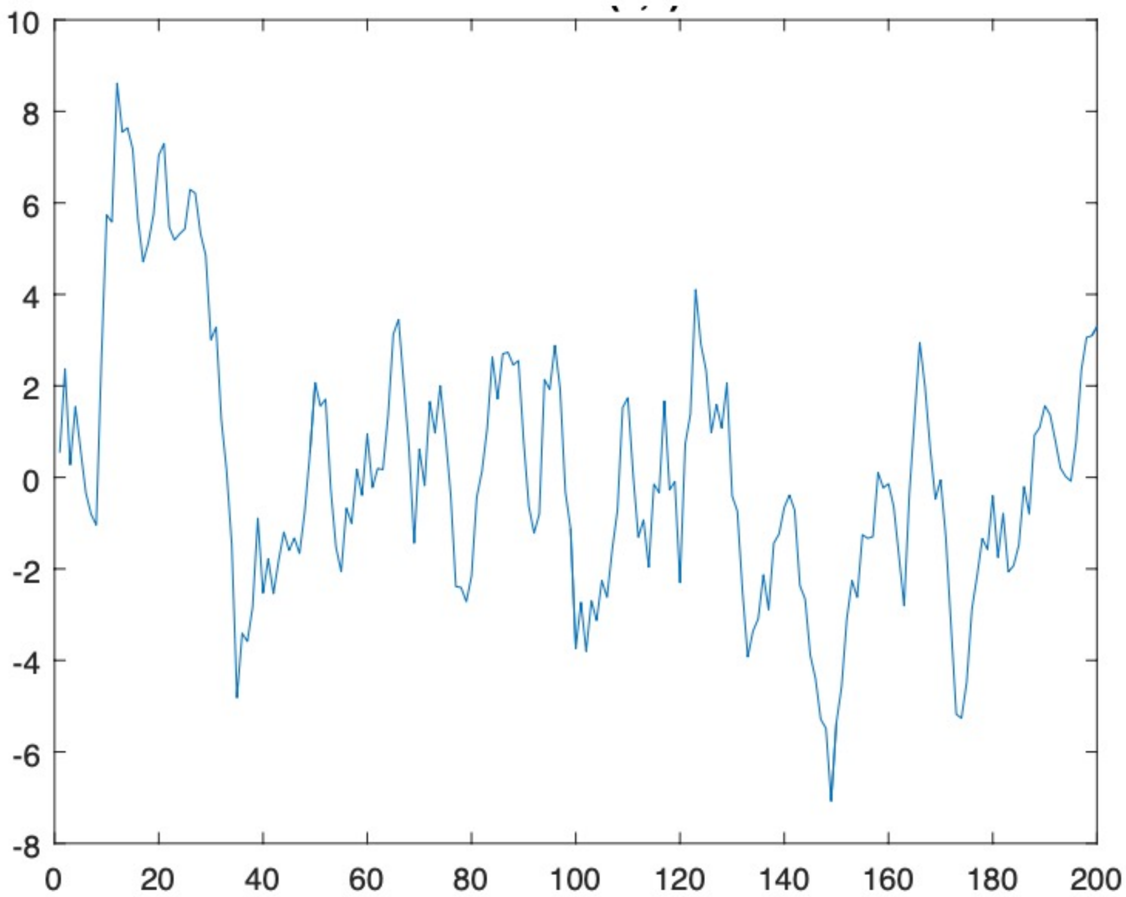
Time Series

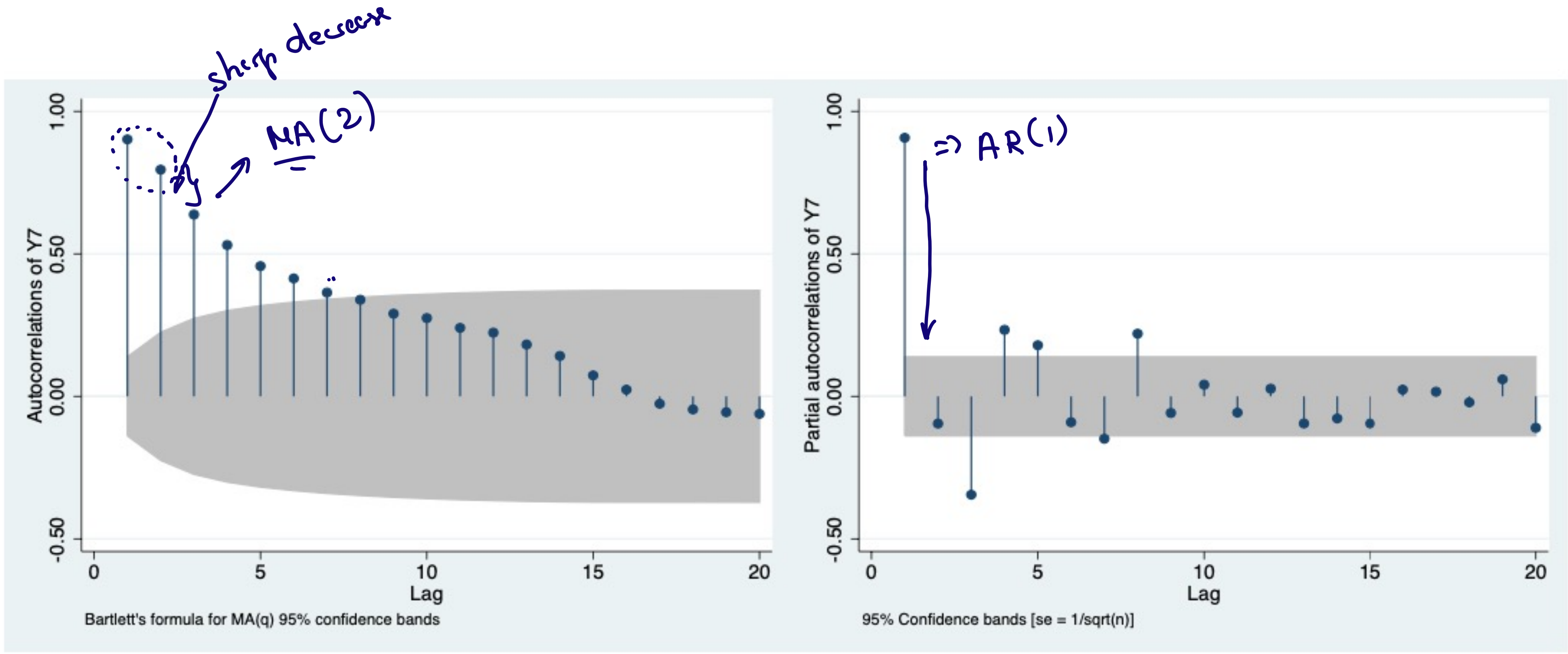
Peter Lukianchenko

24 January 2022

What is this process?



What is this process?



ARMA(1,2): $Y_t = 0.8Y_{t-1} + \varepsilon_t + 0.2\,\varepsilon_{t-1} + 0.5\,\varepsilon_{t-2}$

Non-Stationary time series

- So, an ARMA process

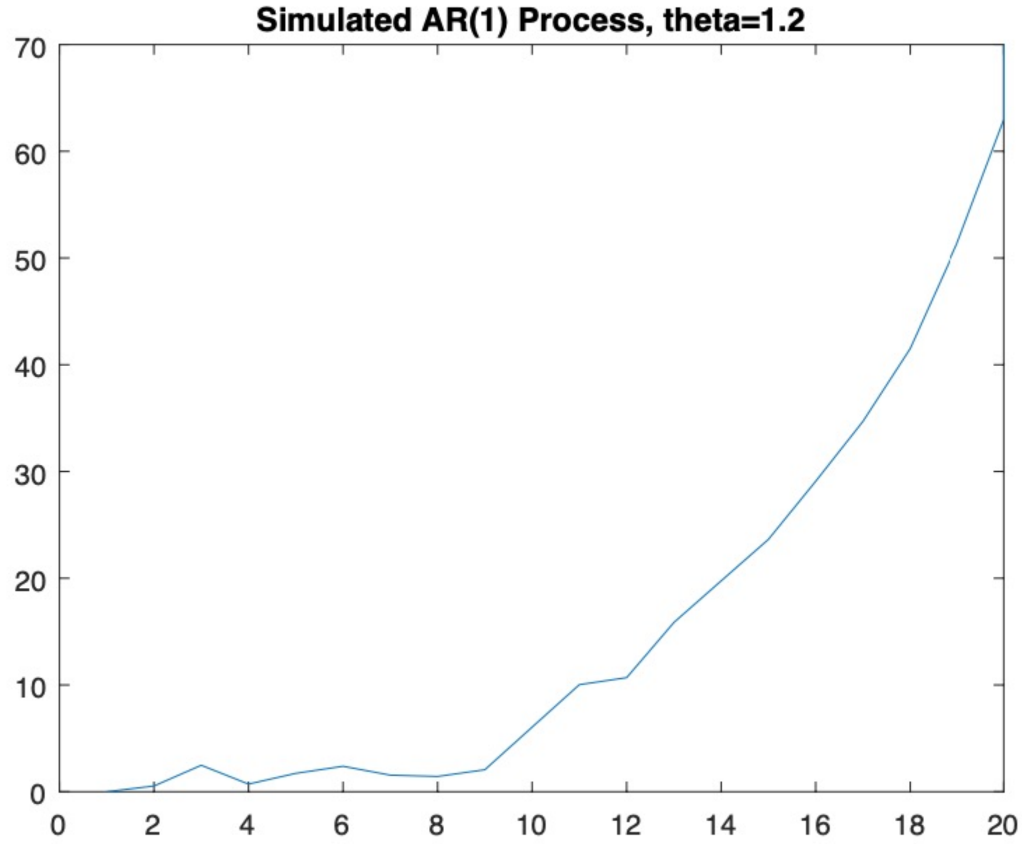
$$\Theta(L)\,Y_t = c + \Phi(L)\,\varepsilon_t$$

  is stationary, if the roots of the characteristic polynomial for $\Theta(L)$ lie outside the unit circle, i.e. if x that solve $\Theta(x) = 0$ are such that $|x| > 1$.

- If a root is inside the unit circle, then the process explodes:

Explosive AR(1)

AR(1) $\quad y_t = 1.3 \cdot y_{t-1} + \varepsilon_t$



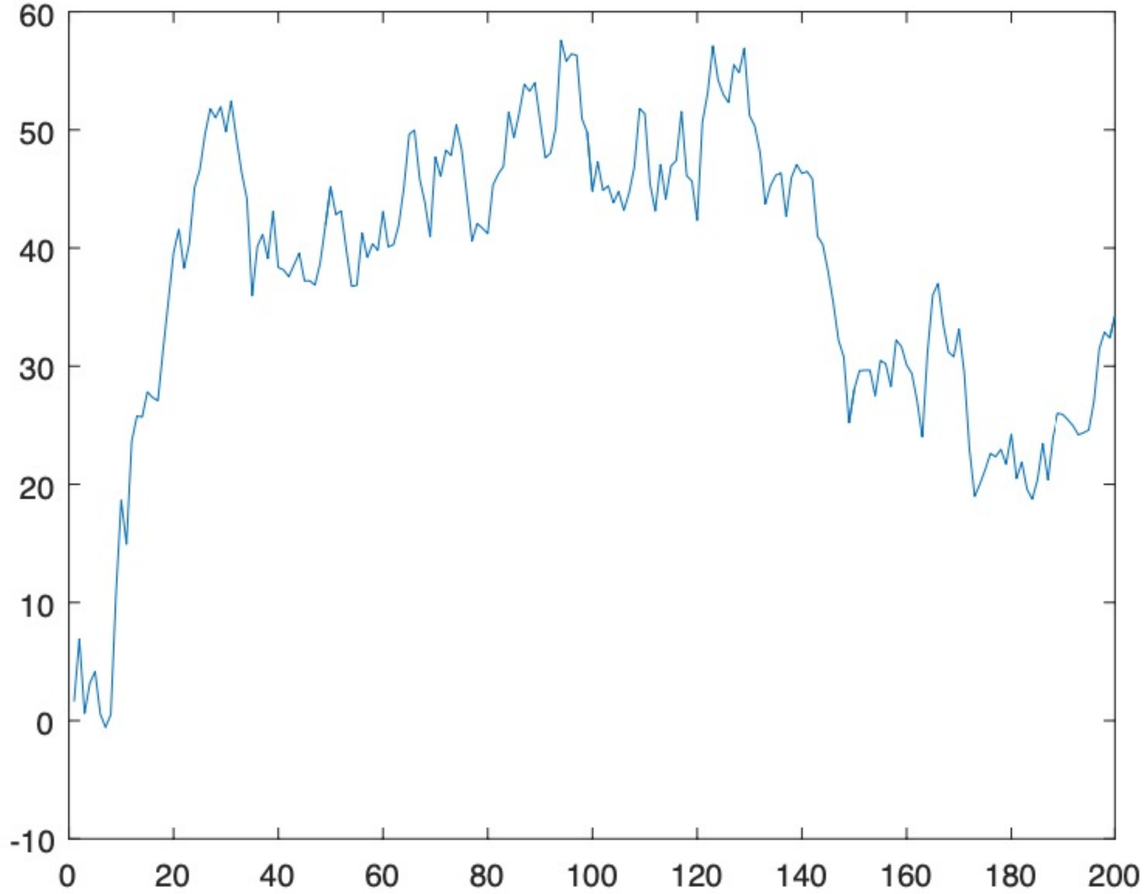
Non-Stationary time series

- Recall: an ARMA process

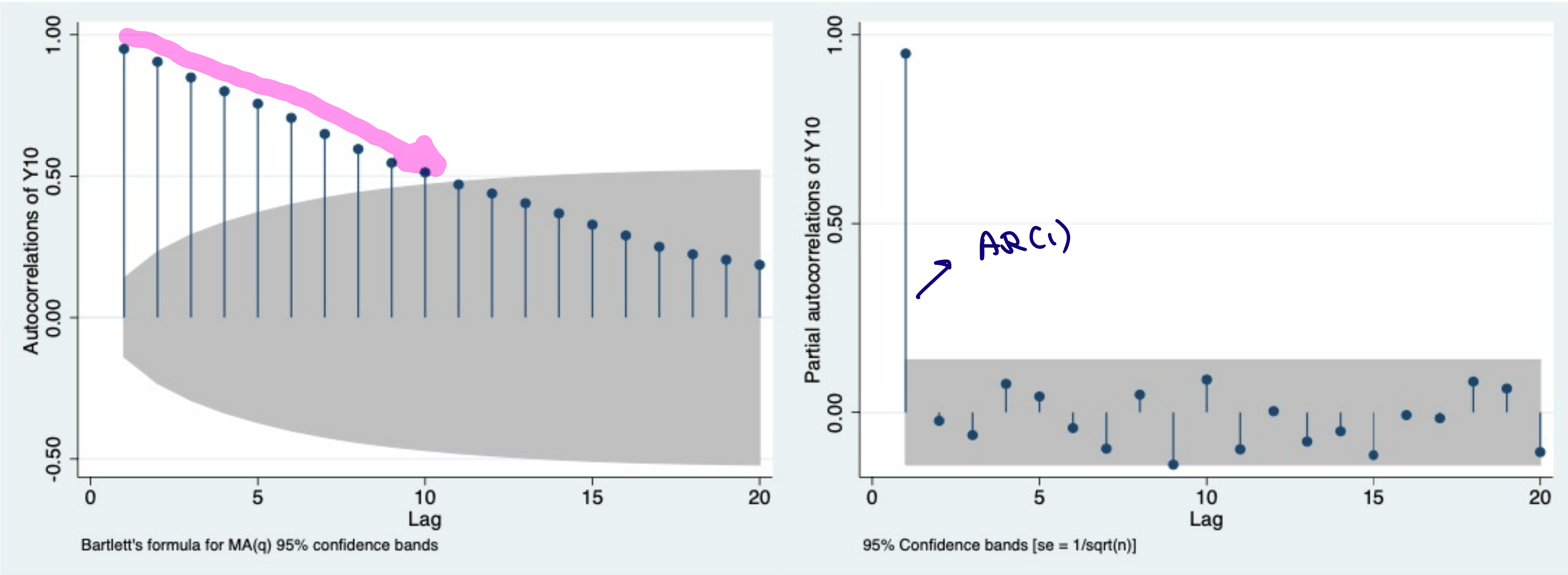
$$\Theta(L)\, Y_t = c + \Phi(L)\, \varepsilon_t$$

  is stationary, if the roots of the characteristic polynomial for $\Theta(L)$ lie outside the unit circle, i.e. if x that solve $\Theta(x) = 0$ are such that $|x| > 1$.

- If a root is inside the unit circle, then the process explodes
- What if a root is **on** the unit circle?


60
50
40
30
20
10
0
-10
0
20
40
60
80
100
120
140
160
180
200

Unit root process



Unit roots

- Many possibilities for a root to be on the unit circle, we care only about $x = 1$, a.k.a. "unit root"

- If there is a unit root, it means that $\{Y_t\}$ can be written as

y_t

$y_t - y_{t-1} = \dots -$

$y_t - y_{t-1} - y_{t-2}$

$$(1-L)^d \Theta(L) Y_t = c + \Phi(L)\,\varepsilon_t,$$

where d is the multiplicity of the unit root, $\Theta(L)$ and $\Phi(L)$ are lag polynomials, with all the roots of $\Theta(L)$ lying outside the unit circle.

- These models are called ***ARIMA(p,d,q)***. p and q are the order of $\Theta(L)$ and $\Phi(L)$.

- $\{Y_t\}$ is also referred to as I(d) – ***integrated of order d***

Integrated process

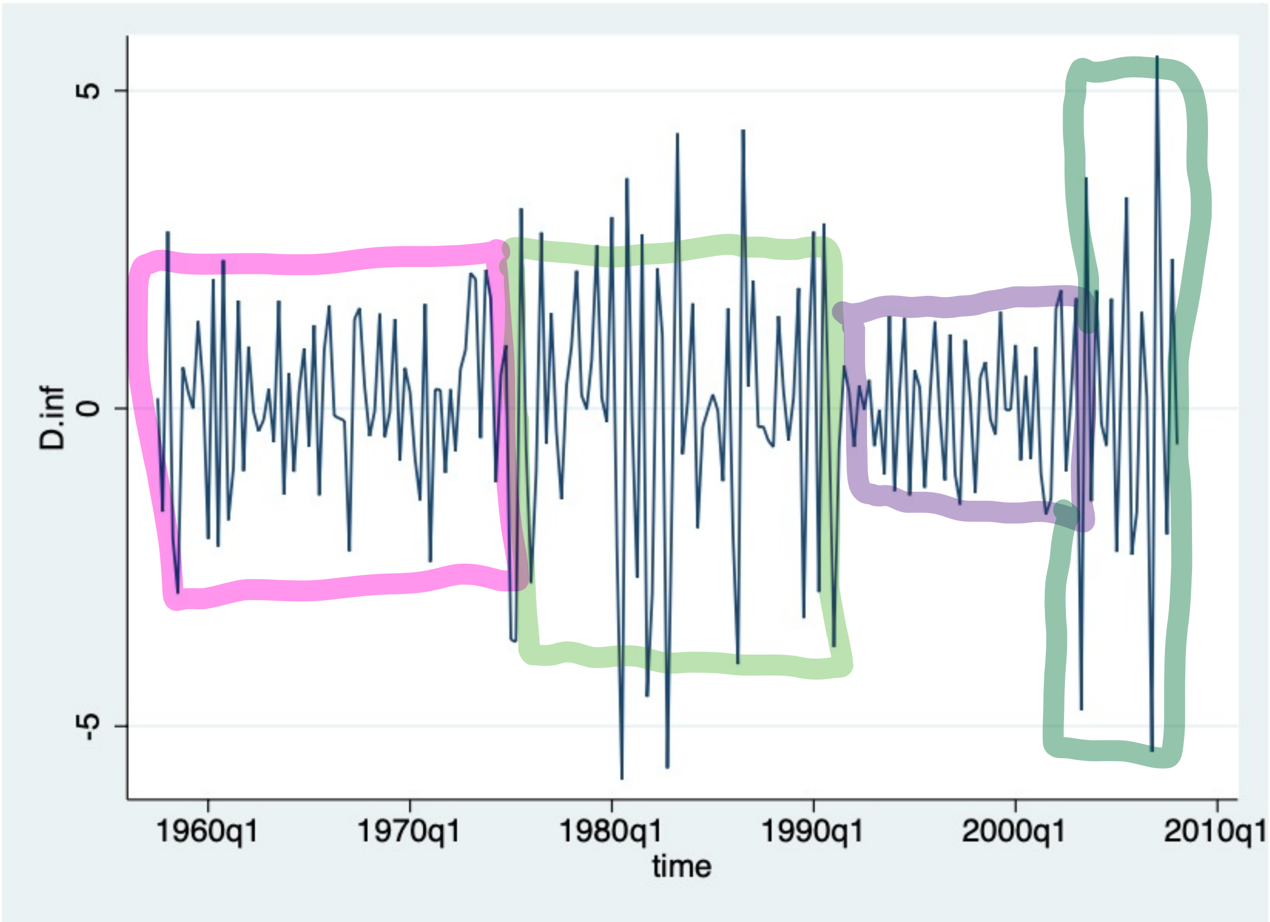
- Define $Z_t = (1-L)^d Y_t$. Then

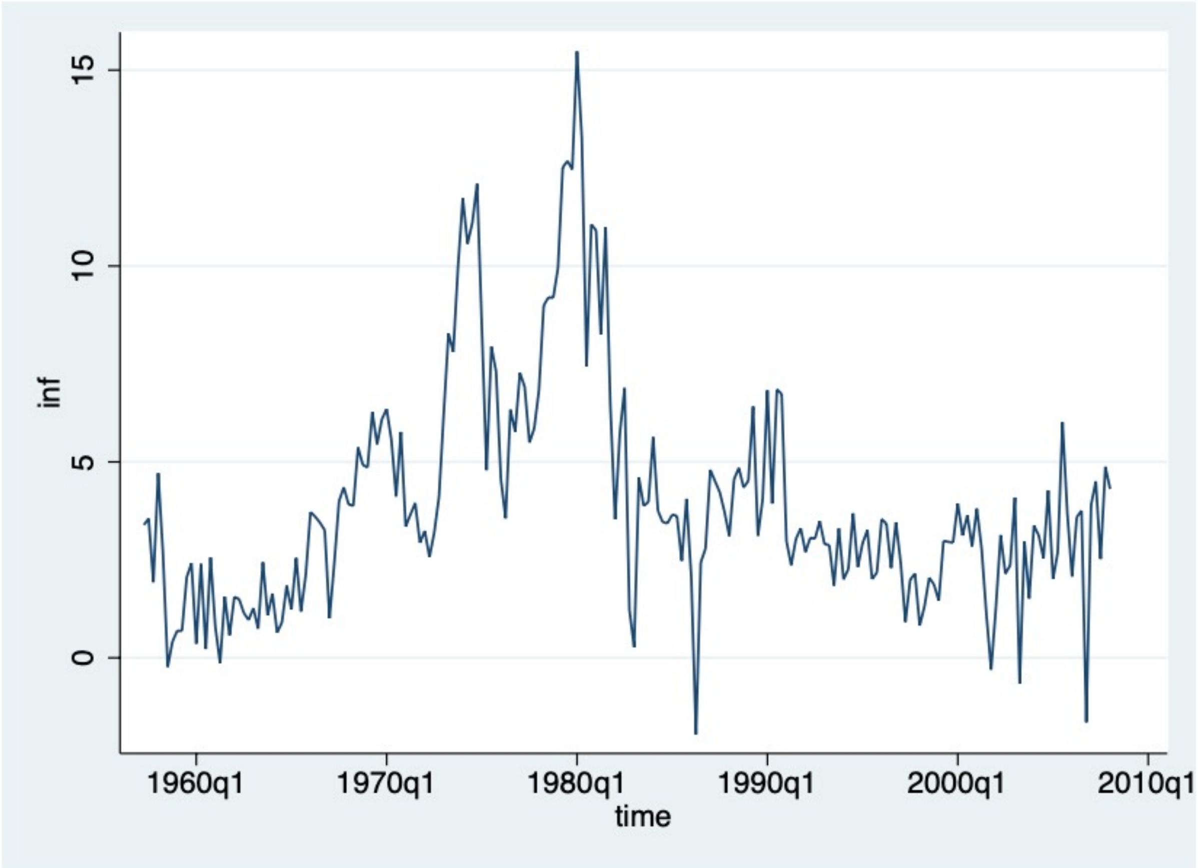
$$\Theta(L) Z_t = c + \Phi(L)\,\varepsilon_t\,.$$

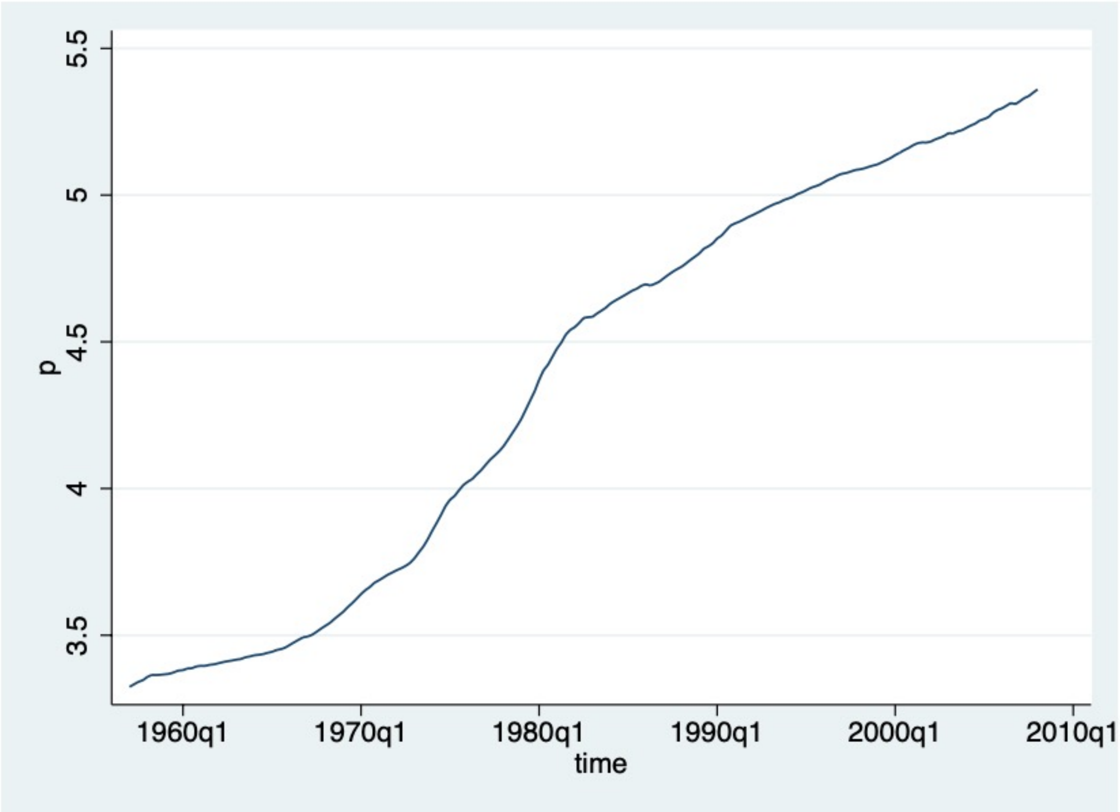
- $\{Z_t\}$ is ARMA(p,q).
- What is $(1-L)^d Y_t$?
  - $d = 1$: $(1-L)\,Y_t = Y_t - Y_{t-1}$ – first difference of Y_t
  - $d = 2$: $(1-L)^2 Y_t = (1-L)(Y_t - Y_{t-1}) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$ – first difference of the first difference of Y_t = second difference of Y_t $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$
  - and so on
- ARIMA(p,d,q) can be made stationary by taking d^{th} difference of Y_t.

ARIMA(p,1,q)

Stationary process




15
10
5
0
inf
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
time


5.5
5
4.5
4
3.5
p
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
time

Integrated process: example

Sometimes from econ theory

Example (Permanent Income Hypothesis, Friedman, 1957)

Consumer solves

$$\max_{\{C_t\}} \mathrm{E}\left[\sum_{t=0}^{\infty} \beta^t u(C_t) | \mathscr{I}_0\right]$$

$$\text{s.t. } A_{t+1} = (1+r)(A_t + Y_t - C_t).$$

If $u(\cdot)$ is quadratic and $\beta = 1/(1+r)$, then

$$C_t = \mathrm{E}_t[C_{t+1}] := \mathrm{E}_t[C_{t+1} | \mathscr{I}_t]$$

Let $\varepsilon_{t+1} = C_{t+1} - C_t$. Then $C_{t+1} = C_t + \varepsilon_{t+1}$ (and $\mathrm{E}_t[\varepsilon_{t+1}] = 0$)

$$C_t = \frac{r}{1+r}\left[A_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j \mathrm{E}_t[Y_{t+k}]\right]$$

Invertability of ARMA

- If AR(p) can be written as MA(∞), perhaps, MA(q) can be written as AR(∞)?
- Yes, if the process is ***invertible***:

Definition

An MA(q) process $\{Y_t\}$ is invertible, if the roots of the characteristic polynomial

$$1 + \varphi_1 x + \varphi_2 x^2 + \ldots + \varphi_q x^q = 0$$

lie outside the unit circle.

To avoid confusing

- We use the word 'invertible' in different context:
  - We say that **the operator $\boldsymbol{\Theta}(\boldsymbol{L})$ is invertible**: this just means that there exists an operator that is inverse to $\Theta(L)$. Denoted by $\Theta^{-1}(L)$. Definition: $\Theta^{-1}(L)\Theta(L) = \Theta(L)\Theta^{-1}(L) = 1$ (identity operator)
  - We say that **an ARMA process $\boldsymbol{Y_t}$ is invertible**, if we can express ε_t as a function of Y_t and its lags

$P(L)$

$y_t = \frac{\varphi(L)}{\theta(L)} \varepsilon_t$

- Don't get confused:
  - An ARMA process $\Theta(L)\,Y_t = \Phi(L)\,\varepsilon_t$ is **stationary**, if **the operator $\boldsymbol{\Theta}(\boldsymbol{L})$** is invertible (i.e., $\boldsymbol{\Theta}^{-1}(\boldsymbol{L})$ exists)
  - An ARMA process $\Theta(L)\,Y_t = \Phi(L)\,\varepsilon_t$ is **invertible**, if **the operator $\boldsymbol{\Phi}(\boldsymbol{L})$** is invertible (i.e., $\boldsymbol{\Phi}^{-1}(\boldsymbol{L})$ exists)

Estimation of ARMA

- AR(p): $Y_t = c + \theta_1 Y_{t-1} + \ldots + \theta_p Y_{t-p} + \varepsilon_t$
- How do you estimate it?
- OLS indeed! Minimize $\sum_{t=p+1}^{T} (Y_t - c - \theta_1 Y_{t-1} - \ldots - \theta_p Y_{t-p})^2$ and get

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

- Here:
  - $\boldsymbol{\theta}$ is a vector with all the coefficients (including c)
  - $\boldsymbol{X}$ is the matrix with observations of the RHS variables:

$$\boldsymbol{X} = \begin{pmatrix} 1 & Y_p & Y_{p-1} & \ldots & Y_2 & Y_1 \\ 1 & Y_{p+1} & Y_p & \ldots & Y_3 & Y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & Y_{T-1} & Y_{T-2} & \ldots & Y_{T-p+1} & Y_{T-p} \end{pmatrix}$$

  - $\boldsymbol{Y} = (Y_{p+1}, Y_{p+2}, \ldots, Y_T)^{\mathrm{T}}$.

Estimation of ARMA

- MA(1):

$$Y_t = \mu + \varepsilon_t + \varphi \varepsilon_{t-1}$$

- How do you estimate it?

RSS

- Can we minimize $\sum_{t=2}^{T} (Y_t - \mu - \varphi \varepsilon_{t-1})^2$?
- Alas, ε_{t-1} is not observed
- Assume $|\varphi| < 1$ and use invertibility of the process

$$\varepsilon_{t-1} = \sum_{j=0}^{+\infty} (-\varphi)^j (Y_{t-j-1} - \mu)$$

Estimation of ARMA models

- MA(1):

$$Y_t = \mu + \varepsilon_t + \varphi \varepsilon_{t-1}$$

-

$$\varepsilon_{t-1} = \sum_{j=0}^{+\infty} (-\varphi)^j (Y_{t-j-1} - \mu)$$

- Can we minimize?

$$\sum_{t=2}^{T} \left[Y_t - \mu - \varphi \sum_{j=0}^{+\infty} (-\varphi)^j (Y_{t-j-1} - \mu) \right]^2$$

- We can minimize (not by hand)

$$\sum_{t=2}^{T} \left[Y_t - \mu - \varphi \sum_{j=0}^{t-2} (-\varphi)^j (Y_{t-j-1} - \mu) \right]^2 .$$

What you should know after today: ML

- Assume ε_t are i.i.d. $\mathcal{N}(0, \sigma^2)$.
- Use invertibility
- Knowing p and q, we can condition on the first several observations on Y and ε and maximize conditional likelihood function
- Not easy to maximize, need to be creative (write ARMA as VAR and use Kalman filter...)
- Works pretty well even if the ε_t are not normal (read about QML)

Box-Jenkins procedure

Step 1
- Look at ACF and PACF
- Get an Idea about p and q (and d)

Step 2
- Estimate the candidate models

Step3
- Compute AIC or BIC, choose the best one
- Do diagnostics

Step 4
- Use the chosen model for forecasting

Choosing the best ARMA

- Hard to tell p and q from the picture
- In general, larger p and q ⇒ better fit. But we like models with smaller p and q
- **Akaike's Information Criterion (AIC):**

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

- **Schwarz's Bayesian Information Criterion (BIC, or SIC)**

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

Choosing the best ARMA

- Choose the model with the smallest IC
- Akaike's Information Criterion (AIC):

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

- Schwarz's Bayesian Information Criterion (BIC, or SIC)

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

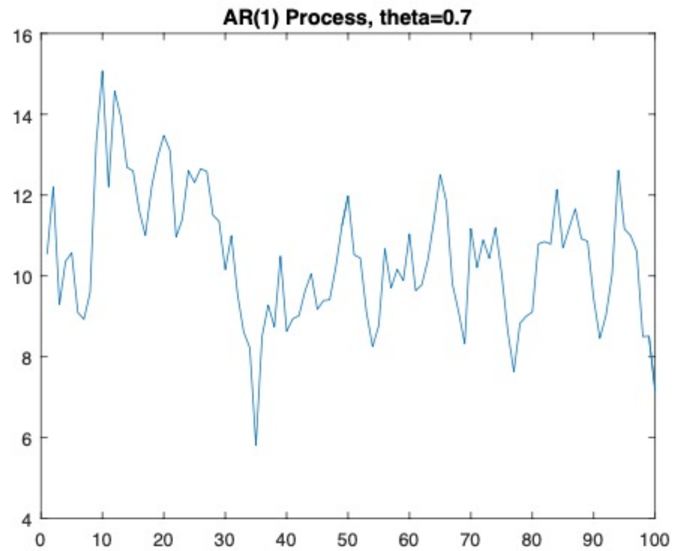
- BIC is better at choosing the correct model asymptotically
- AIC might be better in small samples
- AIC, in general, aims at choosing a model with better forecasting power.

Overview

- Unconditional forecasts: use the marginal distribution of Y_{t+h}, ignore the information at hand
  - Optimal point forecast under square loss: unconditional mean of Y_{t+h}
  - Same for all periods
  - Its risk is equal to $\text{Var}(Y_{t+h}) = \text{Var}(e_{t+h})$
  - Interval forecasts can be formed using the unconditional distribution of Y_{t+h}
- Conditional forecasts: use all the information we have $(Y_t, Y_{t-1}, ...)$
  - Optimal point forecast under square loss: conditional mean of Y_{t+h} given $(Y_t, Y_{t-1}, Y_{t-2}, ...)$
  - Different for different h
  - Its risk is equal to $\text{E}[\text{Var}(Y_{t+h}|Y_t, Y_{t-1}, Y_{t-2}, ...)] = \text{Var}(e_{t+h|t})$
  - Interval forecasts can be formed using the conditional distribution of Y_{t+h}

AR(1)

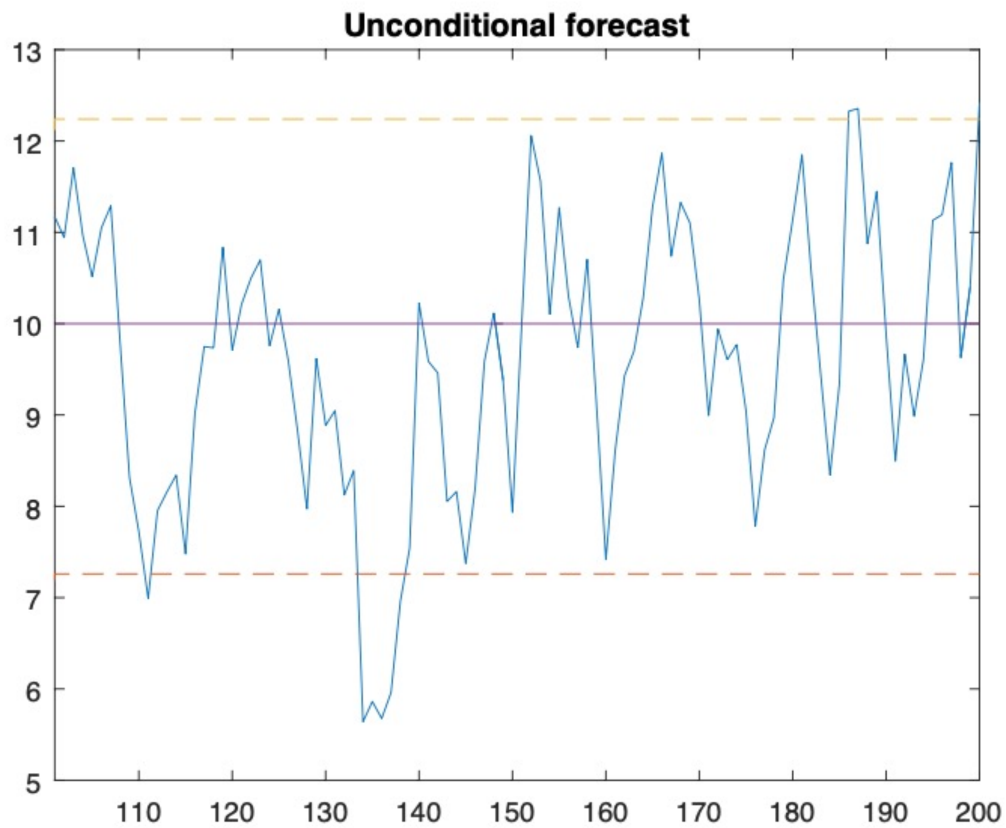
- Let $Y_t = 3 + 0.7Y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}(0,1)$.
- The time series:



Example

- Let $Y_t = 3 + 0.7Y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}(0,1)$.
- $Y_t = 10 + \sum_{j=0}^{+\infty} 0.7^j \varepsilon_{t-j}$
- $Y_t \sim \mathcal{N}(10, 1/0.51)$
- Optimal point forecast: 10
- Optimal 95% interval forecast: $10 \pm 1.96 \cdot \sqrt{1/0.51} \approx [7.26, 12.74]$

Unconditional forecast



Example of [redacted] conditional forecast

- Let $Y_t = 3 + 0.7Y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ are i.i.d. $\mathcal{N}(0,1)$.
- $Y_{100} = 7.16,\ Y_{99} = 8.51, \dots$

$y_{t+h} \to y_{t+h-1} \to y_{t+h-2} \to \dots \to y_{100}$

- Optimal point forecast:

$$\hat{Y}_{t+h|t} = \mathrm{E}[Y_{t+h}|Y_t, Y_{t-1}, \dots] = 3\,\frac{1-0.7^h}{1-0.7} + 0.7^h Y_t$$

- Optimal point forecast:

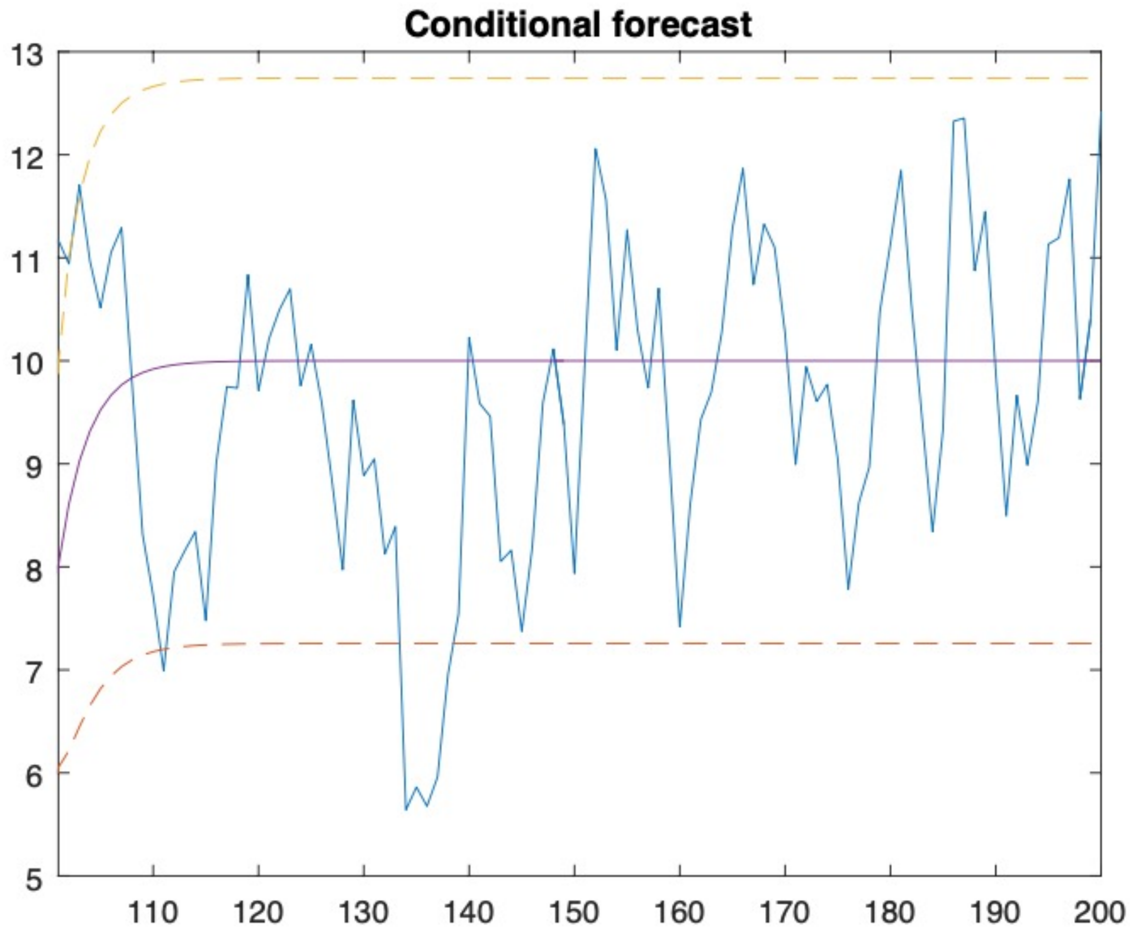
$$\hat{y}_{100+h|100} = \mathrm{E}[Y_{100+h}|Y_{100} = 7.16, Y_{99} = 8.51, \dots] = 3\,\frac{1-0.7^h}{1-0.7} + 0.7^h \cdot 7.16$$

- The conditional distribution of

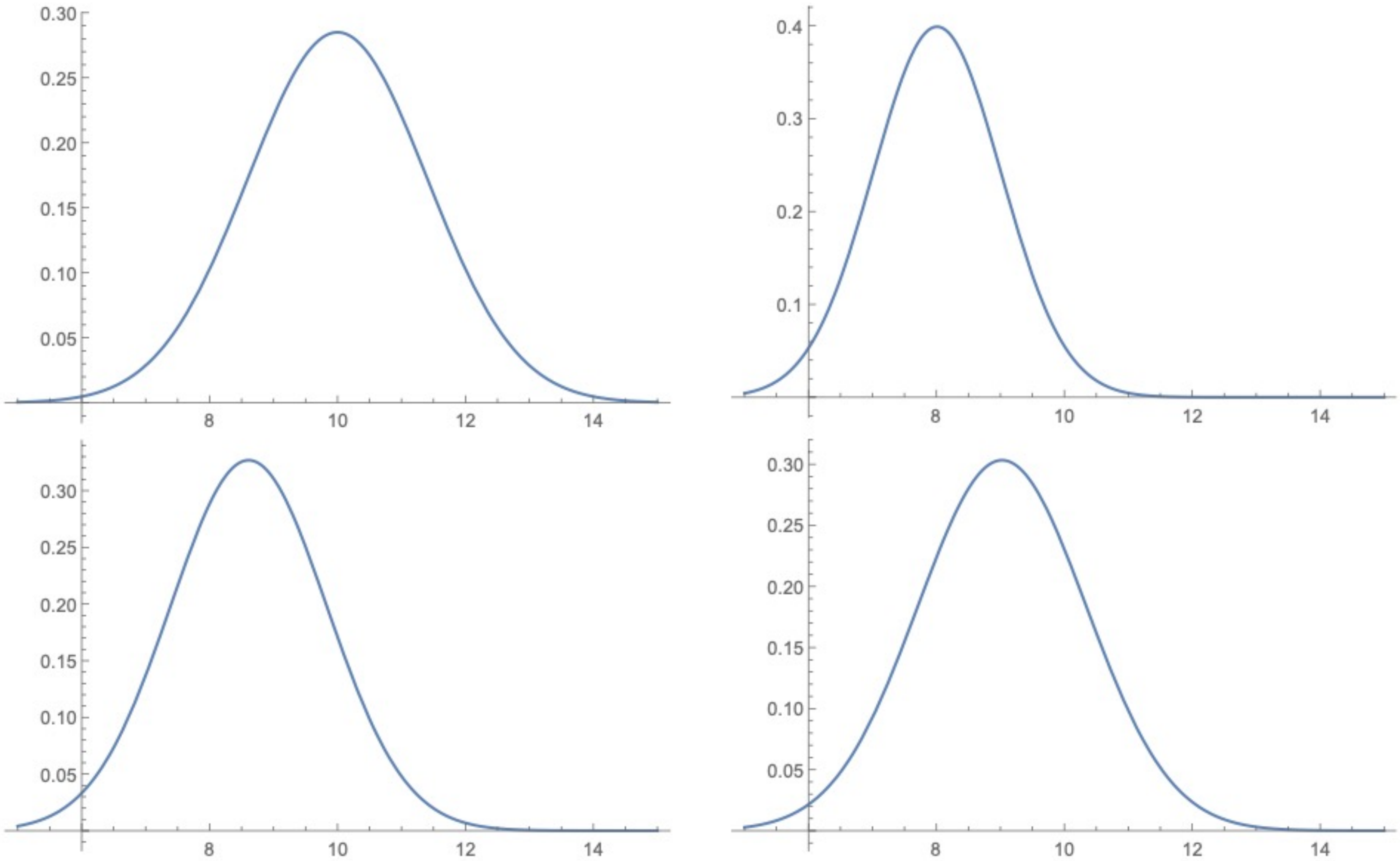
$$Y_{100+h}|Y_{100} = 7.16, Y_{99} = 8.51, \dots \sim \mathcal{N}\left(3\,\frac{1-0.7^h}{1-0.7} + 0.7^h \cdot 7.16,\ \frac{1-\theta^{2h}}{1-\theta^2}\right)$$

- One-period optimal forecast is 8.01; 95% forecast interval is [6.05,9.97]

Conditional forecast



Conditiional forecast



Iterated forecasts

- Consider, for example, AR(1): $Y_t = c + \theta_1 Y_{t-1} + \varepsilon_t$
- Then

$$\hat{Y}_{t+2|t} = \mathrm{E}[Y_{t+2} | Y_t, ...] = \mathrm{E}[c + \theta_1 Y_{t+1} + \varepsilon_{t+2} | Y_t, ...]$$

- So

$$\hat{Y}_{t+2|t} = c + \theta_1 \hat{Y}_{t+1|t}$$

- And in general

$$\hat{Y}_{t+h|t} = c + \theta_1 \hat{Y}_{t+h-1|t}$$

- That's why they are called ***iterated forecasts***
- There are ***direct forecasts*** too

Direct forecasts

- Consider again AR(1): $Y_t = c + \theta_1 Y_{t-1} + \varepsilon_t$
- To forecast 2 periods into the future:

$$Y_{t+2} = c + \theta_1 Y_{t+1} + \varepsilon_{t+2} = c + \theta_1 (c + \theta_1 Y_t + \varepsilon_{t+1}) + \varepsilon_{t+2}$$

- $Y_{t+2} = \alpha + \beta Y_t + u_{t+2}$,

  where $\alpha = c(1+\theta_1)$, $\beta = \theta_1^2$, and $u_t = \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2}$.
- We can simply **estimate the regression** of Y_{t+2} on Y_t
- $\{u_t\}$ is not a white noise anymore: it's MA(1)!
- This is the direct forecast:

$$\hat{Y}^{df}_{t+2|t} = \alpha + \beta Y_t.$$

Direct forecasts

- Consider again AR(1): $Y_t = c + \theta_1 Y_{t-1} + \varepsilon_t$
- We **estimate the regression** of Y_{t+2} on Y_t
- $\{u_t\}$ is not a white noise anymore: it's MA(1)!
- The errors are serially correlated – we need the HAC variance estimator

Direct forecasts

- For the AR(1) example from the beginning:
- $\hat{\alpha} = 5.28$, $\hat{\beta} = 0.49$
- $\hat{y}^{df}_{102|100} = 5.28 + 0.494 \cdot 7.16 = 8.82$
- The iterated forecast was $\hat{y}_{102|100} = 3\frac{1-0.7^2}{1-0.7} + 0.7^2 \cdot 7.16 = 8.61$
- $Y_{102} = 8.31$

Forecast error

- We derived them *knowing the model parameters*
- But in real life we don't know them, we need to estimate them
- We use estimates of the coefficients to compute forecasts
- Now the forecast error also contains the error from the estimation of the coefficients
- Keep that in mind

Comparing models

- Several approaches to determine how good the model is
- Might care about how well the model fits the data
- Or, might care about how good is the predictive ability of the model
- Usually, there is a trade-off between the two

In sample fit

- **In-sample fit:** Estimate the model on existing data, look at an information criterion:

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

- Show how well the model fits the existing sample