

Stat 443. Time Series and Forecasting.

$B$  is the backward shift operator.

Key topics: (a) invertibility for MA polynomial or  $\theta(B)$  in order to get  $h$ -step forecast for ARMA( $p, q$ ),  
(b) invertibility for AR polynomial or  $\phi(B)$  in order to get variance of  $h$ -step forecast for ARMA( $p, q$ )

## Invertibility

Assume  $Y_t = \tilde{Y}_t$  has been centered to have mean 0.  
Consider a stationary ARMA( $p, q$ ) time series model

$$\begin{aligned}Y_t &= \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ \phi(B)Y_t &= \theta(B)\epsilon_t \\ \phi(b) &= 1 - \phi_1 b - \cdots - \phi_p b^p \\ \theta(b) &= 1 + \theta_1 b + \cdots + \theta_q b^q\end{aligned}$$

If  $\phi(b)$  is invertible, then  $(1 - \phi_1 b - \cdots - \phi_p b^p)^{-1} = \phi^{-1}(b)$  is a convergent Taylor series; this is the condition for stationarity.

If  $\theta(b)$  is invertible, then  $(1 + \theta_1 b + \cdots + \theta_q b^q)^{-1} = \theta^{-1}(b)$  is a convergent Taylor series; this is the condition for identifiability (to be explained later via simple cases).

$h$ -step forecasts for ARMA( $p, q$ ):  $\phi(B)(Y_t - \mu) = \theta(B)\epsilon_t$

Invert  $\theta(B)$  to get:

$$\pi(B)(Y_t - \mu) = \theta^{-1}(B)\phi(B)(Y_t - \mu) = \epsilon_t, \quad \pi(b) = 1 - \pi_1 b - \pi_2 b^2 - \dots$$

Suppose we have a data series  $(y_t)$  and estimates  $\hat{\mu}, (\hat{\phi}_i), (\hat{\theta}_k)$  after fitting ARMA( $p, q$ ) leading  $(\hat{\pi}_i)$ .

Shift indices: substitute **forecast values when not yet observed**

$$\begin{aligned} (Y_t - \mu) &= \sum_{i=1}^{\infty} \pi_i (Y_{t-i} - \mu) + \epsilon_t \\ Y_{n+1} &\approx \mu + \sum_{i=1}^n \pi_i (Y_{n+1-i} - \mu) + \epsilon_{n+1} \\ \hat{y}_{n+1|n} &= \hat{\mu} + \sum_{i=1}^n \hat{\pi}_i (y_{n+1-i} - \hat{\mu}) \\ Y_{n+2} &\approx \mu + \sum_{i=1}^{n+1} \pi_i (Y_{n+2-i} - \mu) + \epsilon_{n+2} \\ \hat{y}_{n+2|n} &= \hat{\mu} + \hat{\pi}_1 (\hat{y}_{n+1|n} - \hat{\mu}) + \sum_{i=2}^{n+1} \hat{\pi}_i (y_{n+2-i} - \hat{\mu}) \\ Y_{n+3} &\approx \mu + \sum_{i=1}^{n+2} \pi_i (Y_{n+3-i} - \mu) + \epsilon_{n+3} \\ \hat{y}_{n+3|n} &= \hat{\mu} + \hat{\pi}_1 (\hat{y}_{n+2|n} - \hat{\mu}) + \hat{\pi}_2 (\hat{y}_{n+1|n} - \hat{\mu}) + \sum_{i=3}^{n+2} \hat{\pi}_i (y_{n+3-i} - \hat{\mu}) \\ &\text{etc.} \end{aligned}$$

Special cases of  $p, q$  in later slides

Variance of  $h$ -step forecasts:  $\phi(B)(Y_t - \mu) = \theta(B)\epsilon_t$

Invert  $\phi(B)$  to get:

$$Y_t - \mu = \phi^{-1}(B)\theta(B)\epsilon_t = \psi(B)\epsilon_t, \quad \psi(b) = 1 + \psi_1 b + \psi_2 b^2 + \dots$$

Suppose we have a data series  $(y_t)$  and estimates  $\hat{\mu}, (\hat{\phi}_j), (\hat{\theta}_k)$  after fitting ARMA( $p, q$ ) leading  $(\hat{\psi}_i)$ .

Shift indices:  $\mathcal{F}_n$  is the information to time  $n$ , including  $\epsilon_n = e_n, \epsilon_{n-1} = e_{n-1}, \dots$  when "observed"

$$\begin{aligned} (Y_t - \mu) &= \epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i} \\ Y_{n+1} &\approx \mu + \epsilon_{n+1} + \sum_{i=1}^n \psi_i \epsilon_{n+1-i} \\ [Y_{n+1} | \mathcal{F}_n] &\approx \mu + \epsilon_{n+1} + \sum_{i=1}^n \psi_i e_{n+1-i} \\ \text{Var}(Y_{n+1} | \mathcal{F}_n) &= \text{Var}(\epsilon_{n+1}) = \sigma_\epsilon^2 \\ SE(\hat{y}_{n+1|n}) &= \hat{\sigma}_\epsilon \\ Y_{n+2} &\approx \mu + \epsilon_{n+2} + \psi_1 \epsilon_{n+1} + \sum_{i=2}^n \psi_i \epsilon_{n+2-i} \\ [Y_{n+2} | \mathcal{F}_n] &\approx \mu + \epsilon_{n+2} + \psi_1 \epsilon_{n+1} + \sum_{i=2}^n \psi_i e_{n+2-i} \\ \text{Var}(Y_{n+2} | \mathcal{F}_n) &= \text{Var}(\epsilon_{n+2} + \psi_1 \epsilon_{n+1}) = (1 + \psi_1^2) \sigma_\epsilon^2 \\ SE(\hat{y}_{n+2|n}) &= (1 + \hat{\psi}_1^2)^{1/2} \hat{\sigma}_\epsilon \\ Y_{n+3} &\approx \mu + \epsilon_{n+3} + \psi_1 \epsilon_{n+2} + \psi_2 \epsilon_{n+1} + \sum_{i=3}^n \psi_i \epsilon_{n+3-i} \\ [Y_{n+3} | \mathcal{F}_n] &\approx \mu + \epsilon_{n+3} + \psi_1 \epsilon_{n+2} + \psi_2 \epsilon_{n+1} + \sum_{i=3}^n \psi_i e_{n+3-i} \\ \text{Var}(Y_{n+3} | \mathcal{F}_n) &= \text{Var}(\epsilon_{n+3} + \psi_1 \epsilon_{n+2} + \psi_2 \epsilon_{n+1}) = (1 + \psi_1^2 + \psi_2^2) \sigma_\epsilon^2 \\ SE(\hat{y}_{n+3|n}) &= (1 + \hat{\psi}_1^2 + \hat{\psi}_2^2)^{1/2} \hat{\sigma}_\epsilon \\ SE(\hat{y}_{n+h|n}) &= (1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2)^{1/2} \hat{\sigma}_\epsilon, \quad \text{increasing in } h = 2, \dots \end{aligned}$$

To check with R output in some cases.

Forecasting for stationary AR(2). Training set of length  $n$ .

Let  $\tilde{Y}_t = Y_t - \mu$ .  $\tilde{Y}_i = \phi_1 \tilde{Y}_{i-1} + \phi_2 \tilde{Y}_{i-2} + \epsilon_i$

$$\begin{aligned}
 \tilde{Y}_t &= \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \epsilon_t \\
 &= \phi_1 (\phi_1 \tilde{Y}_{t-2} + \phi_2 \tilde{Y}_{t-3} + \epsilon_{t-1}) + \phi_2 \tilde{Y}_{t-2} + \epsilon_t \\
 &= (\phi_1^2 + \phi_2) \tilde{Y}_{t-2} + \phi_1 \phi_2 \tilde{Y}_{t-3} + \phi_1 \epsilon_{t-1} + \epsilon_t \\
 \tilde{Y}_{n+1} &= \phi_1 \tilde{Y}_n + \phi_2 \tilde{Y}_{n-1} + \epsilon_{n+1} \\
 \hat{y}_{n+1|n} &= \hat{y}_{n+1} = \hat{\mu} + \hat{\phi}_1 (y_n - \hat{\mu}) + \hat{\phi}_2 (y_{n-1} - \hat{\mu}) \\
 \tilde{Y}_{n+2} &= (\phi_1^2 + \phi_2) \tilde{Y}_n + \phi_1 \phi_2 \tilde{Y}_{n-1} + \phi_1 \epsilon_{n+1} + \epsilon_{n+2} \\
 \hat{y}_{n+2|n} &= \hat{\mu} + (\hat{\phi}_1^2 + \hat{\phi}_2) (y_n - \hat{\mu}) + \hat{\phi}_1 \hat{\phi}_2 (y_{n-1} - \hat{\mu}) \\
 &= \hat{y}_{n+2} = \hat{\mu} + \hat{\phi}_1 (\hat{y}_{n+1} - \hat{\mu}) + \hat{\phi}_2 (y_n - \hat{\mu}) \\
 \hat{y}_{n+3|n} &= \hat{y}_{n+3} = \hat{\mu} + \hat{\phi}_1 (\hat{y}_{n+2} - \hat{\mu}) + \hat{\phi}_2 (\hat{y}_{n+1} - \hat{\mu}) \quad \text{etc.}
 \end{aligned}$$

Iterate: use estimated  $\hat{y}_t$  if  $Y_t$  is not observed at time  $n$ .

$$\begin{aligned}
 \text{Var}(Y_{n+1}|\mathcal{F}_n) &= \text{Var}(Y_{n+1}|\text{observations to time } n) = \text{Var}(\epsilon_{n+1}) = \sigma_\epsilon^2 \\
 \text{Var}(Y_{n+2}|\mathcal{F}_n) &= \phi_1^2 \text{Var}(\epsilon_{n+1}) + \text{Var}(\epsilon_{n+2}) = (\phi_1^2 + 1) \sigma_\epsilon^2 \\
 SE(\hat{y}_{n+1|n}) &= \hat{\sigma}_\epsilon \\
 SE(\hat{y}_{n+2|n}) &= (\hat{\phi}_1^2 + 1)^{1/2} \hat{\sigma}_\epsilon
 \end{aligned}$$

Relate to  $\pi(b) = \phi(b)$  and  $\psi(b)$  series.

Variance of forecast for AR(2): use the  $\psi(B)$  series.

$$\begin{aligned}\phi(b) &= 1 - \phi_1 b - \phi_2 b^2 \\ \psi(b) &= \phi^{-1}(b) \text{ because } \theta(b) = 1 \\ 1 &= \psi(b)\phi(b) = (1 + \psi_1 b + \psi_2 b^2 + \psi_3 b^3 + \dots)(1 - \phi_1 b - \phi_2 b^2)\end{aligned}$$

Coefficients for  $b^1, b^2, b^3, \dots$  should be 0 to solve for  $\psi_1, \psi_2, \dots$

$$\begin{aligned}b^1 &: -\phi_1 + \psi_1 = 0 \Rightarrow \psi_1 = \phi_1 \\ b^2 &: -\phi_2 - \psi_1 \phi_1 + \psi_2 = 0 \Rightarrow \psi_2 = \phi_2 + \psi_1 \phi_1 \\ b^3 &: -\psi_1 \phi_2 - \psi_1 \phi_1 + \psi_3 = 0 \Rightarrow \psi_3 = \psi_1 \phi_2 + \psi_2 \phi_1 \\ b^k &: -\psi_{k-2} \phi_2 - \psi_{k-1} \phi_1 + \psi_k = 0 \Rightarrow \psi_k = \psi_{k-2} \phi_2 + \psi_{k-1} \phi_1, \quad k \geq 3\end{aligned}$$

Compare with the `ARMAtoMA()` function in R.

`ARMAtoMA(ar = numeric(), ma = numeric(), lag.max)`

This function can produce the  $\psi_k$  coefficients in the  $MA(\infty)$  representation.

```
phi1 = 0.4; phi2 = 0.2
psiv = rep(0,6)
psiv[1] = phi1; psiv[2] = phi2+psiv[1]*phi1
for(k in 3:6) psiv[k] = psiv[k-2]*phi2 + psiv[k-1]*phi1
psiv2 = ARMAtoMA(ar=c(phi1,phi2), lag.max=6)
print(psiv)
# [1] 0.400000 0.360000 0.224000 0.161600 0.109440 0.076096
print(psiv2)
# [1] 0.400000 0.360000 0.224000 0.161600 0.109440 0.076096

phi1 = 0.4; phi2 = 0.2; theta1 = 0.5
psivec = ARMAtoMA(ar=c(phi1,phi2), ma=c(theta1), lag.max=6)
print(psivec)
# [1] 0.900000 0.560000 0.404000 0.273600 0.190240 0.130816
```

Check output of unempl-tsmode.Rmd

```

dlmy_ar2ml = arima(ytrain[, 'dlmy'], order=c(2,0,0), method="ML")
#>          ar1      ar2    intercept
#>      0.3532 0.1597      -0.0051
#> s.e. 0.1106 0.1120      0.0068      sigma^2 estimated as 0.0009023

phi1 = dlmy_ar2ml$coef[1]; phi2 = dlmy_ar2ml$coef[2]; mu = dlmy_ar2ml$coef[3]
predobj = predict(dlmy_ar2ml, n.ahead=4, se.fit=T)
#>      year      Qtr1      Qtr2      Qtr3      Qtr4
#> $pred 2007 -0.005639112 -0.007769369 -0.006112258 -0.005867239
#> $se   2007  0.03003797  0.03185616  0.03298193  0.03331674

sigma = sigmahat=sqrt(dlmy_ar2ml$sigma2)= 0.03004,
phi1=phi1hat= 0.3532, phi2=phi2hat= 0.1597, mu=muhat= -0.0051
dlmy_ntrain = ytrain[ntrain, 'dlmy'] = -0.02072608
dlmy_nminus1 = ytrain[ntrain-1, 'dlmy'] = 0.02597554

1-step forecast: dlmyhat(n+1) = dlmyhat_nplus1 = -0.005639 =
mu + phi1*(dlmy_ntrain-mu) + phi2*(dlmy_nminus1-mu) =
-0.0051 + 0.3532*(-0.0207261+0.0051) + 0.1597*(0.0259755+0.0051)
2-step forecast: dlmyhat(n+2) = dlmyhat_nplus2 = -0.007769369 =
mu + phi1*(dlmyhat_nplus1-mu) + phi2*(dlmy_ntrain-mu) =
-0.0051 + 0.3532*(-0.005639 +0.0051) + 0.1597*(-0.0207261+0.0051)
SEdlmyhat(n+1) = sigma = 0.03004
SEdlmyhat(n+2) = sigma*sqrt(1+phi1^2) = 0.03186 # psi1=phi1 for AR(2) etc
Exercise: confirm the third and fourth SE in the predict output; see preceding slide

```

## Variance of forecast for AR(1)

$$\tilde{Y}_{n+1} = \phi \tilde{Y}_n + \epsilon_{n+1}$$

$$\tilde{Y}_{n+2} = \phi \tilde{Y}_{n+1} + \epsilon_{n+2} = \phi^2 \tilde{Y}_n + \phi \epsilon_{n+1} + \epsilon_{n+2}$$

$$\tilde{Y}_{n+3} = \phi \tilde{Y}_{n+2} + \epsilon_{n+3} = \phi^3 \tilde{Y}_n + \phi^2 \epsilon_{n+1} + \phi \epsilon_{n+2} + \epsilon_{n+3}$$

$$\tilde{Y}_{n+h} = \phi^h \tilde{Y}_n + \sum_{j=1}^h \phi^{h-j} \epsilon_{n+j}$$

$$\text{Var}(\tilde{Y}_{n+h} | Y_n = y_n, \dots, Y_1 = y_1) = \sigma_\epsilon^2 \sum_{j=1}^h \phi^{2(h-j)}$$

$$SE(\hat{y}_{n+h|n}) = \hat{\sigma}_\epsilon \left[ \sum_{j=1}^h \hat{\phi}^{2(h-j)} \right]^{1/2}$$

Confirm with AR(1) model for furnace data in webwork exercise

ARMA(1,1): Deriving  $\pi(b)$  and  $\psi(b)$ 

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \quad -1 < \phi < 1, \quad -1 \leq \theta \leq 1.$$

$$(1 - \phi B)Y_t = (1 + \theta B)\epsilon_t, \quad \text{what happens if } \phi = -\theta?$$

Deducing  $\psi(b)$

- $\psi(b) = (1 - \phi b)^{-1}(1 + \theta b) = (1 + \psi_1 b + \psi_2 b^2 + \dots)$
- $(1 + \psi_1 b + \psi_2 b^2 + \dots)(1 - \phi b) = (1 + \theta b)$
- Match powers of  $b$
- $b^1$ :  $-\phi + \psi_1 = \theta \Rightarrow \psi_1 = \theta + \phi$
- $b^2$ :  $-\psi_1 \phi + \psi_2 = 0 \Rightarrow \psi_2 = \phi \psi_1$
- $b^k$ :  $-\psi_{k-1} \phi + \psi_k = 0 \Rightarrow \psi_k = \phi \psi_{k-1} = \phi^{k-1} \psi_1, \quad k \geq 2.$

Deducing  $\pi(b)$

- $\pi(b) = (1 + \theta b)^{-1}(1 - \phi b) = (1 - \pi_1 b - \pi_2 b^2 - \dots)$
- $(1 - \pi_1 b - \pi_2 b^2 - \dots)(1 + \theta b) = (1 - \phi b)$
- Match powers of  $b$
- $b^1$ :  $\theta - \pi_1 = -\phi \Rightarrow \pi_1 = \theta + \phi$
- $b^2$ :  $-\pi_1 \theta - \pi_2 = 0 \Rightarrow \pi_2 = -\theta \pi_1$
- $b^k$ :  $-\pi_{k-1} \theta - \pi_k = 0 \Rightarrow \pi_k = -\theta \pi_{k-1} = (-\theta)^{k-1} \pi_1, \quad k \geq 2.$

Use similar steps for ARMA(2,1)

## Take-home message

1. When using statistical method for the first time in a software, check that you can match each outputted value to theory.
2. ARMA( $p, q$ ) with  $q \geq 1$ : 1-step forecast  $\hat{y}_{t+1|t}$  is a linear function of  $y_t, \dots, y_1$  but can be truncated to the most recent  $m$  previous observations where the coefficients are not negligible.