MSA 8200 Predictive Analytics
Week 4: ARMA model and alternatives

Yichen Cheng

Spring 2020

Outline

- ARMA model
- Forcasting
- Estimation
- Model Diagnostics
- Alternatives?

ARMA model – Definition

A time series $\{x_t; t = 0, \pm 1, \ldots\}$ is ARMA(p,q) if:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

- $\phi_p \neq 0$, $\theta_q \neq 0$, $\sigma_w^2 > 0$.
- Stationary.
- p is the AR order and q is the MA order.

ARMA model – Definition

Definition: AR and MA polynomials:

- AR polynomials: $\Phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$, $\phi_p \neq 0$.
- MA polynomials: $\Theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$, $\theta_q \neq 0$.
- z is a complex number.

ARMA model – Redundancy

Parameter redundancy:

$$\Phi(B)x_t = \Theta(B)w_t$$
$$\Rightarrow \eta(B)\Phi(B)x_t = \eta(B)\Theta(B)w_t$$

Example:

- Ture model: $x_t = w_t$.
- Let $\eta(B) = 1 - 0.5B$, then we have

$$(1 - 0.5B)x_t = (1 - 0.5B)w_t$$
$$\Rightarrow x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$$

- It looks like an ARMA model.

ARMA model – Redundancy

We use ARMA(p,q) model to refer to its simplest form.

- That is $\Phi(z)$ and $\theta(z)$ have no common factors.

$$x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t.$$

- $\Phi(z) = 1 - 0.5z$.
- $\Theta(z) = 1 - 0.5z$.
- $\Phi(z)$ and $\Theta(z)$ have common factor.
- $\Phi(z) = 1$ and $\Theta(z) = 1$.
- It is a ARMA(0,0) process.

ARMA model – Causality

Definition 3.7: Concept of causality

- An ARMA(p,q) model for x_t is said to be causal if x_t can be written as a one-sided linear process:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \Psi(B) w_t$$

- where $\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.
- $\psi_0 = 1$.
- It is also called the MA representation.

ARMA model – Causality

An ARMA(p,q) model is causal if and only if:

- $\Phi(z) \neq 0$ for $|z| \leq 1$.
- Then $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$.
- $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \Theta(z)/\Phi(z)$, $|z| \leq 1$.
- Equivalently $\Phi(z) = 0$ only when $|z| > 1$.

ARMA model – Invertibility

Definition 3.8: An ARMA(p,q) model is invertible, if the time series $\{x_t : t = 0, \pm 1, \pm 2, \dots\}$ can be written as

$$\Pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t$$

- $\Pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$;
- $\sum_{j=0}^{\infty} |\pi_j| < \infty$;
- $\pi_0 = 1$;
- It is called the invertible representation.

ARMA model – Invertibility

Prop 3.2: Invertibility of an ARMA(p,q) process.

An ARMA(p,q) model is invertible if and only if

- $\Theta(z) \neq 0$ for $|z| \leq 1$.
- Then, $\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \Phi(z)/\Theta(z)$, $|z| \leq 1$.
- In other words, roots of $\Theta(z)$ lie outside the unit circle.
- $\Theta(z) = 0$ only when $|z| > 1$.

ARMA model – Example

Example 3.7: Parameter redundancy, causality and invertibility

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

- $(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t$.
- It appears to be an ARMA(2,2) process.

- $\Phi(B) = (1 - 0.4B - 0.45B^2) = (1 - 0.9B)(1 + 0.5B)$.
- $\Theta(B) = (1 + B + 0.25B^2) = (1 + 0.5B)^2$.
- Common factor: $(1 + 0.5B)$. ARMA(1,1)!

ARMA model – Example

Exemple 3.7 continued:

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t$$

- $x_t = 0.9x_t + 0.5w_{t-1} + w_t$.

- Causal: $\Phi(z) = (1 - 0.9z) = 0 \Rightarrow z = 10/9$.
- Invertible: $\Theta(z) = (1 + 0.5z) = 0 \Rightarrow z = -2$.

ARMA model – Example

Exemple 3.7 continued:

$$(1 - 0.9B)x_t = (1 + 0.5B)w_t$$

- The MA representation:

$$x_t = (1 + \psi_1 B + \psi_2 B^2 + \ldots)w_t.$$

- $(1 - 0.9z)(1 + \psi_1 z + \psi_2 z^2 + \ldots) = 1 + 0.5z$
- $1+(\psi_1 - 0.9)z+(\psi_2 - 0.9\psi_1)z^2+\ldots+(\psi_j - 0.9\psi_{j-1})z^2+\ldots = 1+0.5z.$

- $\psi_1 - 0.9 = 0.5 \Rightarrow \psi_1 = 1.4.$
- $\psi_j - 0.9\psi_{j-1} = 0 \Rightarrow \psi_j = 1.4 \times 0.9^{(j-1)}$

How about the invertible representation?

Forecasting

Goal for forecasting:

- Use current and past observations to predict future.

Forecasting

Goal for forecasting:

- Use current and past observations to predict future.

- Predict x_{n+1} based on $X_{1:n} = \{x_1, x_2, \ldots, x_n\}$.
- Predict x_{n+2} based on $X_{1:n} = \{x_1, x_2, \ldots, x_n\}$.
- $\ldots$

Forecasting

Goal: predict $x_{n+m}, m = 1, 2, \ldots$ based on $X_{1:n} = \{x_1, x_2, \ldots, x_n\}$

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$$

- x_{n+m}^n minimize $E[x_{n+m} - x_{n+m}^n]^2$,
- It is called the best linear predictor. (BLP)

Forecasting

Property 3.3: Best linear prediction for stationary process

- The parameters $\alpha_0, \alpha_1, \ldots, \alpha_n$ in $x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k$ is found by solving

$$E[(x_{n+m} - x_{n+m}^n)x_k] = 0, k = 0, 1, \ldots, n,$$

for $\alpha_0, \alpha_1, \ldots, \alpha_n$, where $x_0 = 1$.

Forecasting

One-step-ahead prediction:

$$x_{n+1}^{n} = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \ldots + \phi_{nn}x_1 = \underline{\phi}_n'\underline{x} \tag{1}$$

- $E[(x_{n+1} - \sum_{j=1}^{n} \phi_{nj}x_{n+1-j})x_{n+1-k}] = 0, k = 1, \ldots, n.$
  $\Rightarrow \sum_{j=1}^{n} \phi_{nj}\gamma(k-j) = \gamma(k)$ or $\Gamma_n\underline{\phi}_n = \underline{\gamma}_n$.

$$\underline{\phi}_n = \Gamma_n^{-1}\underline{\gamma}_n. \tag{2}$$

- $\Gamma_n = \{\gamma(k-j)\}, j = 1, \ldots, n, k = 1, \ldots, n$
- $\underline{\gamma}_n = (\gamma(1), \ldots, \gamma(n))'$

Forecasting

Mean squared one-step-ahead prediction error:

$$P_{n+1}^{n} = E(x_{n+1} - x_{n+1}^{n})^2 = \gamma(0) - \underline{\gamma}_n' \Gamma_n^{-1} \underline{\gamma}_n.$$

- $\Gamma_n = \{\gamma(k-j)\}, j = 1, \dots, n, k = 1, \dots, n$
- $\underline{\gamma}_n = (\gamma(1), \dots, \gamma(n))'$

Forecasting

Example Prediction for an AR(1) model

$$x_t = 0.5x_{t-1} + w_t$$

One-step-ahead prediction of x_2 based on x_1 is:

- $x_2^1 = \phi_{11} x_1$
- $\phi_{11} = \Gamma_1^{-1} \gamma_1$
- $\Gamma_1 = \gamma(0) = \sigma_w^2/(1-\phi^2)$
- $\gamma_1 = \gamma(1) = \sigma_w^2 \phi/(1-\phi^2)$

$$\Rightarrow x_2^1 = 0.5x_1$$

- The mean squared prediction error (MSPE):

$$P_2^1 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n = \gamma(0) - \gamma_1^2/\gamma(0) = \sigma_w^2$$

Example 3.19 Prediction for an AR(2) model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

One-step-ahead prediction of x_2 based on x_1:

$$x_2^1 = \phi_{11} x_1 = \frac{\gamma(1)}{\gamma(0)} x_1 = \rho(1) x_1$$

One-step-ahead prediction of x_3 based on x_2 is:

$$x_3^2 = \phi_{21} x_2 + \phi_{22} x_1$$

- $\phi_{21} = \phi_1$, $\phi_{22} = \phi_2$.

Forecasting – Durbin-Levinson Algorithm

- For general ARMA models, the prediction equations not as simple.
- Also, for n large, using Equation (2) is hard (matrix inverse is expensive).

Solve the prediction iteratively through Durbin-Levinson Algorithm:

- $\phi_{00} = 0$, $P_1^0 = \gamma(0)$
- for $n \geq 1$,
  - $\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}$
  - $P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2)$
- for $n \geq 2$: $\phi_{nk} = \phi_{n-1,k} - \phi nn \phi_{n-1,n-k}$, $k = 1, 2, \ldots, n-1$

Note: the Durbin-Levinson Algorithm can be used to find the PACF. (Example 3.21 in textbook.)

Forecasting

Example 3.25 Forecasting the Recruitment Series.

- Estimation
- Forecasting
- Mean sqaure prediction error (MSPE)

Estimation

Example 3.29 Method of moments estimation for an MA(1)

$$x_t = w_t + \theta w_{t-1}$$

with $|\theta| < 1$.
Use the same idea as for Yule-Walker Equation:

- $\gamma(0) = \sigma_w^2(1 + \theta^2)$
- $\gamma(0) = \sigma_w^2 \theta$

$\hat{\theta}$ can be found by solving

$$\hat{\rho}(1) = \frac{\hat{\gamma}(0)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1 + \hat{\theta}^2} \tag{3}$$

Estimation

Example 3.29 Method of moments estimation for an MA(1)

$$x_t = w_t + \theta w_{t-1}$$

with $|\theta| < 1$.
$\hat{\theta}$ can be found by solving

$$\hat{\rho}(1) = \frac{\hat{\gamma}(0)}{\hat{\gamma}(0)} = \frac{\hat{\theta}}{1 + \hat{\theta}^2} \tag{4}$$

- Two solutions exist, pick the invertible one.
- If $|\hat{\rho}(1)| \leq 1/2$, the solutions are real. Otherwise, no real solution exists.
- Example: $\hat{\rho}(1) = .507$ while $\rho(1) = .9/(1 + .9^2) = .497$.

Estimation – Method of moments

Example 3.29 Method of moments estimation for an MA(1)

- When $|\hat{\rho}(1)| \leq 1/2$, the invertible estimate is

$$\hat{\theta} = \frac{1 - \sqrt{4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$$

Furthermore, it can be shown that

$$\hat{\theta} \sim AN\left(\theta, \frac{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8}{n(1-\theta^2)^2}\right)$$

AN means asymptotically normal.

- In contrast, the maximum likelihood estimator has an asymptotic variance $(1-\theta^2)/n$.
- When $\theta = 0.5$, the ratio is about 3.5.

Estimation – Maximum Likelihood Estimation

AR(1): $x_t = \phi x_{t-1} + w_t$ with $|\phi| < 1$ and $w_t \sim iid\mathcal{N}(0, \sigma_w^2)$

Likelihood:

$$L(\phi, \sigma_w^2) = f(x_1)f(x_2|x_1)\dots f(x_n|x_{n-1})$$

- $x_t|x_{t-1} \sim \mathcal{N}(\phi x_{t-1}, \sigma_w^2)$
  $f(x_t|x_{t-1}) = f_w(x_t - \phi x_{t-1})$, f_w is the density of w_t
- $x_1 = w_1 + \phi w_0 + \phi^2 w_{-1} + \dots = \sum_{j=0}^{\infty} \phi^j w_{1-j}$
  $x_1 \sim \mathcal{N}(0, \sigma_w^2/(1-\phi^2))$

Likelihood:

$$L(\phi, \sigma_w^2) = (2\phi\sigma_w^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left\{-\frac{S(\phi)}{2\sigma_w^2}\right\}$$

- $S(\phi) = (1-\phi^2)x_1^2 + \sum_{t=2}^{n}(x_t - \phi x_{t-1})^2$

Estimation – Maximum Likelihood Estimation

For a general ARMA model:

- It is difficult to write the likelihood as an explicit function of the parameters.
- Instead, it is advantageous to write it in terms of the innovations, one-step-ahead prediction erros $x_t - x_t^{t-1}$

Let $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$:

$$L(\beta, \sigma_w^2) = \prod_{t=1}^{n} f(x_t | x_{t-1}, \dots, x_1)$$

We have $x_t | x_{t-1}, \dots, x_1 \sim \mathcal{N}(x_t^{t-1}, P_t^{t-1})$, where

- $P_t^{t-1} = \gamma(0) \prod_{j=1}^{t-1} (1 - \phi_{jj}^2)$
- $\gamma(0) = \sigma_w^2 \prod_{j=0}^{\infty} \psi_j^2$

The Newton-Raphson algorithm:

- Start with an initial guess β^0.
- At iteration t: update the estimate by':

$$\beta^t = \beta^{t-1} + \{l''(\beta^{-t1})\}^{-1} l'(\beta^{t-1}) \tag{5}$$

- Stop until convergence.

Large Sample Property:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma_w^2 \Gamma_{p,q}^{-1}) \tag{6}$$

Estimation – Overfitting Caveat

Example 3.35 Overfitting Caveat.
Does it matter?

- If the time series follows an AR(1) process.
- We decide to fit an AR(2) model to the data.

By overfitting:

- We loose efficiency:
- AR(1): $var(\hat{\phi}_1) \approx (1 - \phi_1^2)/n$
- fit an AR(2): $var(\hat{\phi}_1) \approx (1 - \phi_2^2)/n = 1/n > (1 - \phi_1^2)/n$

Side note: it can be used as a diagnostic tool.

Example 3.33: Fitting the glacial varve series

- Paleoclimatic glacial varve thicknesses from Massachusetts for $n = 634$ years.
- Melting glaciers deposit yearly layers of sand and silt during the spring melting seasons.
- Begining 11,834 years ago.

To do:

- Estimation.

Model diagnostics

Looking at the residual:

- The residuals should be i.i.d $\mathcal{N}(0, \sigma^2)$

This suggests we shoud check whether the residuals:

- follow a normal distribution. Q-Q plot
- are independent: no auto correlation. Ljung-Box-Pierce Q-statistics

Model diagnostics

Standardized Residuals::

$$e_t = (x_t - x_t^{t-1})/\sqrt{P_t^{t-1}}$$

Ljung-Box-Pierce Q-statistics:

$$Q = n(n+2)\sum_{h=1}^{H}\frac{\hat{\rho}_e^2(h)}{n-h}$$

- $\hat{\rho}_e^2(h)$: lag h auto correlation for $\{e_1, \ldots, e_n\}$
- $Q\ \chi^2_{H-p-q}$.

Model diagnostics – Examples

Example 3.40: Diagnostics for GNP growth rate example
Example 3.41: Diagnostics for the glacial varve series

Basic steps for fitting an ARMA model

Basic steps for fitting ARMA models to time series data:

- plotting the data
- possibly transforming the data
- identifying the dependence orders of the model.
- parameter estimation
- diagnostics
- model choice

Go back to the stock prediction example

Alternatives? https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

- Moving Average
- Linear Regression
- k-Nearest Neighbors (KNN)
- Auto ARIMA
- Prophet
- Long Short Term Memory (LSTM)

What was done right/wrong?

Reading Materials

- Ch 3.3 p.83-88 (ARMA model, Causality, Invertibility)
- Ch 3.4 (Forecasting)
- Ch 3.5 (Estimation)
- Ch 3.7 p.135-142 (Building ARIMA models)
- A short YouTube video on Ljung Box test: https://www.youtube.com/watch?v=FETjCDysdWY