STATS 207: Time Series Analysis Autumn 2020

Lecture 7: ARMA/ARIMA Modelling

Dr. Alon Kipnis

October 5th 2020

Genera Info

- Homework assignment 1 is due today.
- Late submission policy: -20pt for each late submission day up to three late submission days.
- Homework assignment 2 will be out shortly. Due Monday 10/19/2020.
- Guest lecture next week: David Donoho.

Outline

Forecasting

Estimating **ARMA** parameters

Forecasting

Forecasting – So Far

- We seek for the **best linear predictor** of x_{n+m} based on $x_1, \ldots, x_n$

$$x_{n+m}^{n,*} = \sum_{k=1}^{n} \alpha_{n+m,k}^{n,*} x_k = \boldsymbol{\alpha}_{n+1}^{'n,*} x.$$

- The **vector of optimal weights** satisfies

$$\boldsymbol{\alpha}_{n+1}^{n,*} = \boldsymbol{\Gamma}_n^{-1} \gamma_{n+m}^{n}.$$

Forecasting III – Under Stationarity

- If x_t is stationary

$$(\boldsymbol{\Gamma}_n)_{ij} = \mathbb{E}\left[x_i x_j\right] = \gamma_x(i-j), \qquad \text{(Toeplitz, non-negative definite)}$$

$$(\boldsymbol{\gamma}^n_{n+m})_k = \mathbb{E}\left[x_{n+m} x_k\right] = \gamma_x(n+m-k).$$

- Example: One-step ahead prediction

$$\boldsymbol{\Gamma}_n \boldsymbol{\alpha}^{n,*}_{n+1} = \boldsymbol{\gamma}^n_{n+1}.$$

Prediction error satisfies

$$P^n_{n+1} \equiv \mathbb{E}\left[(x_{n+1} - x^1_{n+1})^2\right] = \gamma_x(0) - \boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}_n \boldsymbol{\gamma}^n_{n+1}$$

Proof:

$$\begin{aligned} P^n_{n+1} &= \mathbb{E}\left[(x_{n+1} - \boldsymbol{\alpha}'^{n,*}_{n+1} x)^2\right] = \mathbb{E}\left[(x_{n+1} - \boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n x)^2\right] \\ &= \mathbb{E}\left[x^2_{n+1} - 2\boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n x x_{n+1} + \boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n x x' \boldsymbol{\Gamma}^{-1}_n \boldsymbol{\gamma}^n_{n+1}\right] \\ &= \gamma(0) - 2\boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n \boldsymbol{\gamma}^n_{n+1} + \boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n \boldsymbol{\Gamma}_n \boldsymbol{\Gamma}^{-1}_n \boldsymbol{\gamma}^n_{n+1} \\ &= \gamma(0) - \boldsymbol{\gamma}'^n_{n+1} \boldsymbol{\Gamma}^{-1}_n \boldsymbol{\gamma}^n_{n+1}. \end{aligned}$$

- The **Durbin-Levinson Algorithm** (Property 3.2) solves $\boldsymbol{\alpha}^{n,*}_{n+1}$ and P^n_{n+1} recursively.

One-Step Ahead Forecasting for AR(p)

- Example 3.19: One-step ahead forecasting with causal **AR**(2):

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

  - Based on $x_1, x_2, \ldots$

$$\boldsymbol{\alpha}^{n,*}_{n+1,n} = \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}^n_{n+1},$$

$$\alpha^{n,*}_{n+1,n} = \phi_1, \qquad \alpha^{n,*}_{n+1,n-1} = \phi_2, \qquad \alpha^{n,*}_{n+1,k} = 0, \quad k \geq n-2.$$

- Obvious generalization to **AR**(p).
- Asymptotic relation to π weights

$$\lim_{n\to\infty} \alpha^{n,*}_{n+1,n-k} = -\pi_k, \qquad k = 1, 2, \ldots, n$$

m-step Ahead Forecasting

- m-step ahead prediction error satisfies

$$P_{n+m}^n \equiv \mathbb{E}\left[(x_{n+m} - x_{n+m}^n)^2\right] = \gamma(0) - \boldsymbol{\gamma}_{n+m}^{\prime n}\boldsymbol{\Gamma}_n\boldsymbol{\gamma}_{n+m}^n.$$

- **Prediction intervals**:

$$x_{n+m}^n \pm c_{\alpha/2}\sqrt{P_{n+m}^n}, \qquad \Pr(|w_t| \geq c_{\alpha/2}) \leq \alpha$$

(need adjustment for multiple testing to get PI for more than one time period).

- Example: For **AR**(1) we have

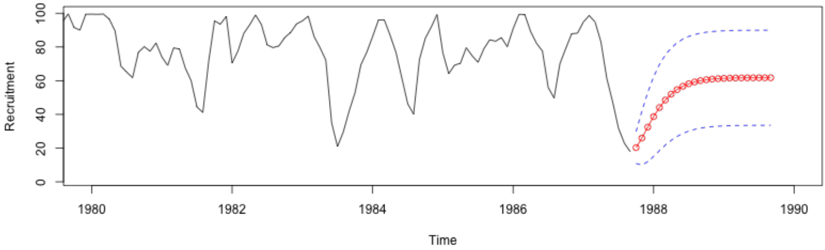
$$P_{n+1}^n = \sigma_w^2.$$

- Example: Long-range prediction

$$\lim_{m\to\infty} P_{n+m}^n = \gamma_x(0).$$

Forecasting – Example 3.26

```
regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE)
fore = predict(regr, n.ahead=24)
ts.plot(rec, fore$pred, col=1:2, xlim=c(1980,1990), ylab="Recruitment")
lines(fore$pred, type="p", col=2)
lines(fore$pred+fore$se, lty="dashed", col=4)
lines(fore$pred-fore$se, lty="dashed", col=4)
```



Estimating ARMA parameters

Estimation – Motivation

- For an **ARMA**(p, q) model with **all parameters given**, we know:
  - Theoretical properties (invertibility, causality).
  - Description of second moments (ACF and PACF).
  - Predict future observations.

- Next:
  - **Estimate ARMA(p, q) parameters** given p and q under invertibility and causality assumptions.
  - Later: How to find p and q?
  - Later: Is **ARMA** even a good model?

Parameter Estimation in AR(p)

Three techniques:

1. Method of moments (MoM) and Yule-Walker estimates
2. Conditional Least Squares (C-LS)
3. Maximum Likelihood (ML)

MoM I: Estimating the Mean

- Use
$$\hat{\mu} = \bar{x} \equiv \frac{1}{n} \sum_{j=1}^{n} x_j.$$
- Theorem A.5: If x_t is a linear process and $\sum_j \psi_j \neq 0$, then
$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} \mathcal{N}(0, V), \quad V = \sum_{h=-\infty}^{\infty} \gamma_x(h) = \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$
- Henceforth assume $\mu = 0$.

MoM II: Estimating AR(p) Parameters Yule-Walker Equations

- Definition: The **Yule-Walker equations** are

$$\gamma(h) = \phi_1\gamma(h-1) + \ldots + \phi_p\gamma(h-p), \qquad h = 1, 2, \ldots, p,$$
$$\sigma_w^2 = \gamma(0) - \phi_1\gamma(1) - \ldots - \phi_p\gamma(p).$$

- In matrix notation

$$\boldsymbol{\Gamma}_p\boldsymbol{\phi} = \boldsymbol{\gamma}_p, \quad \sigma_w^2 = \gamma(0) - \boldsymbol{\phi}'\boldsymbol{\gamma}_p,$$

$$\boldsymbol{\Gamma}_p = \{\gamma(k-j)\}_{j,k=1}^p, \qquad \boldsymbol{\phi} = (\phi_1, \ldots, \phi_p), \qquad \boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))'$$

- **Method of moments** estimation (Yule-Walker estimators): Solve

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p, \qquad \hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\gamma}}_p'\hat{\boldsymbol{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p.$$

where

$$\hat{\boldsymbol{\Gamma}}_p = \{\hat{\gamma}(k-j)\}_{j,k=1}^p, \qquad \hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))'.$$

- We can calculate $\hat{\phi}$ without inverting $\hat{\boldsymbol{\Gamma}}_p$ using the **Durbin-Levinson algorithm.**

Yule-Walker Estimators – Large Sample Result

- Property 3.8: For a causal **AR**(p) process, as $n \to \infty$,

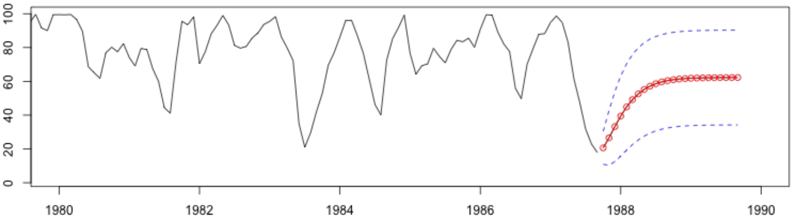
$$\sqrt{n}\left(\hat{\phi} - \phi\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_w^2 \mathbf{\Gamma}_p^{-1}\right), \qquad \hat{\sigma}_w^2 \xrightarrow{P} \sigma_w^2.$$

YW Estimation of the Recruitment Series

```
rec.yw = ar.yw(rec, order=2)
rec.yw$x.mean  # = 62.26278 (mean estimate)
rec.yw$ar      # = 1.3315874, -.4445447  (parameter estimates)
sqrt(diag(rec.yw$asy.var.coef))  # = .04222637, .04222637  (standard errors
rec.yw$var.pred  # = 94.79912 (error variance estimate)
```

Predicting using estimated **AR**(2) parameters:

```
rec.pr = predict(rec.yw, n.ahead=24)
U = rec.pr$pred + rec.pr$se
L = rec.pr$pred - rec.pr$se
minx = min(rec,L); maxx = max(rec,U)
ts.plot(rec, rec.pr$pred, xlim=c(1980,1990), ylim=c(minx,maxx))
lines(rec.pr$pred, col="red", type="o")
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
```



MoM III: Method of Moments Estimation for MA(1)

Example 3.29:

- Consider

$$x_t = w_t + \theta w_{t-1}, \quad |\theta| < 1.$$

- Recall $\gamma_x(0) = \sigma_w^2(1+\theta^2)$ and $\gamma_x(1) = \sigma_w^2\theta$. An estimator $\hat{\theta}$ for θ is obtained from

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

  Two solutions exist. **We pick the invertible one**.

- If $|\hat{\rho}_x(1)| \leq 1/2$,

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

  and (using the Delta Method)

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{(1-\theta^2)^2}\right).$$

- If $|\hat{\rho}_x(1)| > 1/2$, **a real solution does not exist**.
- **Later**: The ML estimator has a better large sample behavior.

ML Estimation

- Definition: **Maximum Likelihood (ML) Estimator**
  1. Write likelihood function in terms of model parameters

     $$L(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2) = L(x_1, \ldots, x_n; \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2)$$

  2. Solve

     $$\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_q, \hat{\sigma}_w^2 \equiv \text{argmin}\, L(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_w^2).$$

Example: (conditional) ML Estimation of AR(1)

$$x_t = \phi x_{t-1} + w_t, \qquad |\phi| < 1, \qquad w_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_w^2).$$

- Likelihood function (conditioned on x_1)

$$L(\phi, \sigma_w^2) = f(x_1, \ldots, x_n; \phi, \sigma_w^2) = \prod_{j=2}^{n} f(x_j | x_{j-1}; \phi, \sigma_w^2),$$

where

$$f(x_j | x_{j-1}; \phi, \sigma_w^2) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{ -\left(x_j - \phi x_{j-1}\right)^2 / \sigma_w^2 \right\}, \qquad j = 2, \ldots, n$$

- ML estimate

$$\hat{\phi} = \mathrm{argmin}_{\phi<1} S_c(\phi; x_1), \qquad \hat{\sigma}_w^2 = S_c(\hat{\phi}; x_1)/(n-1),$$

where

$$S_c(\phi; x_1) \equiv \sum_{j=2}^{n} (x_j - \phi x_{j-1})^2.$$

Conditional LS Estimate

- Definition: **Conditional Least Squares Estimator for ARMA**(p, q)
  1. Set $\hat{w}_t = 0$ for $t \leq p$.
  2. Write **LS objective**

$$\hat{w}_t = x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} - \theta_1 \hat{w}_{t-1} - \ldots - \theta_q \hat{w}_{t-q}, \quad t = p+1, \ldots, n.$$

$$S_c(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) \equiv \sum_{t=p+1}^{n} \hat{w}_t^2.$$

  3. Solve linear regression

$$\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q = \text{argmin}\, S_c(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q).$$

  4. Set

$$\hat{\sigma}_w^2 \equiv \frac{1}{n - p - 1} S_c(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$$

- For **AR**(1), C-LS is almost identical to ML.

Example: Fitting AR(2) to Recruitment

Example 3.31

```
rec.mle = ar.mle(rec, order=2) rec.mle$x.mean
rec.mle$ar
sqrt(diag(rec.mle$asy.var.coef))
rec.mle$var.pred
```

```
62.2615261054892
1.35128085590502 -0.461273619173842
0.0409915944162095 0.0409915944162095
89.3359654276631
```

Compare with YW equations (Example 3.28):

```
rec.yw = ar.yw(rec, order=2)
rec.yw$x.mean
rec.yw$ar
sqrt(diag(rec.yw$asy.var.coef))
rec.yw$var.pred
```

```
62.2627816777042
1.33158738866791 -0.444544697634474
0.0422263743755033 0.0422263743755033
94.7991188417802
```

Fitting AR(2) to Recruitment (cont'd)

Compare with C-LS (Example 3.18):

```
fit <- ar.ols(rec , order=2, demean=TRUE , intercept=TRUE)
fit$x.mean
fit$ar
fit$asy.se.coef$ar
fit$var.pred
```

```
62.2627816777042
1.35406847266143 -0.46317843167489
0.041789006654309 0.0418794219793695
89.71705
```

Large Sample Distribution

Property 3.10: Consider a causal invertible **ARMA**(p, q) process. Denote

$$\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_p).$$

Under appropriate conditions, C-LS estimate and ML estimate, all provide **asymptotically optimal** estimates $\hat{\beta}$ of β and $\hat{\sigma}_w^2$ of σ_w^2. In particular

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_w^2 \mathbf{\Gamma}_{p,q}^{-1}\right), \qquad \mathbf{\Gamma}_{p,q} \equiv \begin{pmatrix} \mathbf{\Gamma}_{\phi\phi} & \mathbf{\Gamma}_{\phi\theta} \\ \mathbf{\Gamma}_{\theta\phi} & \mathbf{\Gamma}_{\theta\theta} \end{pmatrix},$$

Where:

- $\mathbf{\Gamma}_{\phi\phi} \equiv \{\gamma_x(i-j)\}_{i,j=1}^{p}$ where $\gamma_x(h)$ is the autocovariance of **AR**(p) $\phi(B)x_t = w_t$
- $\mathbf{\Gamma}_{\theta\theta} \equiv \{\gamma_y(i-j)\}_{i,j=1}^{q}$ where $\gamma_y(h)$ is the autocovariance of **MA**(q) $y_t = \theta(B)w_t$.
- $\mathbf{\Gamma}_{\phi\theta} \equiv \{\gamma_{xy}(i-j)\}_{i,j=1}^{p,q}$ where $\gamma_{xy}(h)$ is the crosscovariance of x_t and y_t.
- $\mathbf{\Gamma}_{\theta\phi} = \mathbf{\Gamma}'_{\phi\theta}$.

Specific Asymptotic Distributions

- **AR**(1):

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, 1 - \phi^2).$$

- **AR**(2):

$$\sqrt{n}\left(\begin{pmatrix}\hat{\phi}_1 \\ \hat{\phi}_2\end{pmatrix} - \begin{pmatrix}\phi_1 \\ \phi_2\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1 - \phi_2^2 \end{pmatrix}\right)$$

- **MA**(1):

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1 - \theta^2)$$

- **ARMA**(1, 1):

$$\sqrt{n}\left(\begin{pmatrix}\hat{\phi} \\ \hat{\theta}\end{pmatrix} - \begin{pmatrix}\phi \\ \theta\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} (1-\phi^2)^{-1} & (1+\phi\theta)^{-1} \\ (1+\phi\theta)^{-1} & (1-\theta^2)^{-1} \end{pmatrix}^{-1}\right)$$

Note: Variance inflation due to fitting **AR**(2) to **AR**(1).

Recap

- **ARMA**(p, q) is a useful model for **stationary processes**.
- We can express **ACF** and **optimal linear** m**-step forecast** in terms of model's parameters.
- We can **fit ARMA**(p, q) **to data** using several techniques. All techniques lead to asymptotically normal estimators.

Next:

- Extensions (ARIMA, SARIMA)
- How to select p and q?
- Is **ARMA** a good model for my data?