Time Series lecture 4

Parametric estimation

Sofia Olhede

EPFL

March 12, 2025

Lecture outline

1. Yule-Walker (method-of-moments) for AR models

2. Least squares for Gaussian mean-zero AR models

3. Least squares for general ARMA models

Yule-Walker (method-of-moments) for AR models

Setup

For the first two sections of this lecture, assume that X_t is a mean-zero causal AR process of order p:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$
$$= \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_t,$$

where $\epsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ is a Gaussian white noise process.
We suppose that we have observed $X_1, \ldots, X_n$ and we want to estimate the parameters $\phi_1, \ldots, \phi_p$ and σ_ϵ^2.

Yule-Walker derivation

To derive the estimation method we start by multiplying through the AR equation by X_{t-k} for $k > 0$

$$X_t X_{t-k} = \sum_{i=1}^{p} \phi_i X_{t-i} X_{t-k} + \epsilon_t X_{t-k}$$

$$\Rightarrow \mathbb{E}\left[X_t X_{t-k}\right] = \sum_{i=1}^{p} \phi_i \mathbb{E}\left[X_{t-i} X_{t-k}\right] + \mathbb{E}\left[\epsilon_t X_{t-k}\right]$$

- Remember the process is causal, so $\mathbb{E}\left[\epsilon_t X_{t-k}\right] = 0$.
- Furthermore, we have assumed X_t is mean-zero, so

$$\mathbb{E}\left[X_{t-j} X_{t-k}\right] = \gamma_{k-j}.$$

- Therefore

$$\gamma_k = \sum_{j=1}^{p} \phi_j \gamma_{k-j} \tag{4.1}$$

Therefore using (4.1) and symmetry of $\gamma_\tau = \gamma_{-\tau}$

$$\begin{aligned}
\gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 + \cdots + \phi_p\gamma_{p-1}, \\
\gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 + \cdots + \phi_p\gamma_{p-2}, \\
&\vdots \\
\gamma_p &= \phi_1\gamma_{p-1} + \phi_2\gamma_{p-2} + \cdots + \phi_p\gamma_0.
\end{aligned} \tag{4.2}$$

- In matrix form we have

$$\boldsymbol{\gamma} = \boldsymbol{\Gamma}\boldsymbol{\phi}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^T$, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$ and

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}.$$

- If the mean is zero take

$$\hat{\gamma}_\tau = \frac{1}{n}\sum_{t=1}^{n-|\tau|} X_t X_{t+|\tau|}$$

- We estimate ϕ via

$$\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\gamma}}$$

- Now we just need to estimate σ_ϵ^2. We have

$$\mathbb{E}\left[X_t^2\right] = \phi_1 \mathbb{E}\left[X_t X_{t-1}\right] + \cdots + \phi_p \mathbb{E}\left[X_t X_{t-p}\right] + \mathbb{E}\left[\epsilon_t X_t\right].$$

- Therefore

$$\gamma_0 = \sum_{j=1}^{p} \phi_j \gamma_j + \mathbb{E}\left[\epsilon_t X_t\right]$$

- Furthermore,

$$\epsilon_t X_t = \phi_1 \epsilon_t X_{t-1} + \cdots + \phi_p \epsilon_t X_{t-p} + \epsilon_t^2$$

- We take expectations of the left and right hand side and get that

$$\begin{aligned}\mathbb{E}\left[\epsilon_t X_t\right] &= \phi_1 \mathbb{E}\left[\epsilon_t X_{t-1}\right] + \cdots + \phi_p \mathbb{E}\left[\epsilon_t X_{t-p}\right] + \mathbb{E}\left[\epsilon_t^2\right] \\ &= \phi_1 \cdot 0 + \cdots + \phi_p \cdot 0 + \sigma_\epsilon^2 \\ \Rightarrow \gamma_0 &= \sum_{j=1}^{p} \phi_j \gamma_j + \sigma_\epsilon^2.\end{aligned}$$

- Thus the estimator is $\widehat{\sigma}_\epsilon^2 = \hat{\gamma}_0 - \sum_{j=1}^{p} \widehat{\phi}_j \hat{\gamma}_j$.

Definition 4.1 (Yule-Walker estimators)

The estimators $\widehat{\phi}_p$ and $\widehat{\sigma}_\epsilon^2$ are the Yule-Walker estimators.

$$\widehat{\boldsymbol{\phi}}_p = \widehat{\boldsymbol{\Gamma}}_p^{-1} \widehat{\boldsymbol{\gamma}}_p \tag{4.3}$$

$$\widehat{\sigma}_\epsilon^2 = \hat{\gamma}_0 - \sum_{j=1}^{p} \widehat{\phi}_j \hat{\gamma}_j \tag{4.4}$$

- Inverting $\boldsymbol{\Gamma}_p$ naively is $O\left(p^3\right)$, but the Levinson-Durbin algorithm is $O\left(p^2\right)$ by using the structure of $\boldsymbol{\Gamma}_p$ (symmetric Toeplitz matrix).

Least squares for Gaussian mean-zero AR models

Forward least squares: formulation

In order to formulate a least squares approach we begin from the definition of an AR process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t.$$

Specifically, we can write

$$\begin{aligned} X_{p+1} &= \phi_1 X_p + \phi_2 X_{p-1} + \cdots + \phi_p X_1 + \epsilon_{p+1} \\ &\vdots \\ X_n &= \phi_1 X_{n-1} + \phi_2 X_{n-2} + \cdots + \phi_p X_{n-p} + \epsilon_n \end{aligned} \tag{4.5}$$

- Other terms $X_1, \ldots, X_p$ have to be discarded, as we do not observe times before X_1.

Forward least squares: matrix form

We can rewrite this in matrix form

$$\mathbf{X}_F = \mathbf{F}\boldsymbol{\phi} + \boldsymbol{\epsilon}$$

where

$$\mathbf{X}_F = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_n \end{pmatrix}, \quad \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{p+1} \\ \vdots \\ \epsilon_n \end{pmatrix},$$

and

$$\mathbf{F} = \begin{pmatrix} X_p & X_{p-1} & \cdots & X_1 \\ X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \ddots & \dots \\ X_{n-1} & X_{n-2} & \cdots & X_{n-p} \end{pmatrix}.$$

Forward least squares: objective function

We can estimate ϕ by finding the vector which minimises

$$\begin{aligned}\mathrm{SS}(\phi) &= \|\mathbf{X}_F - \mathbf{F}\phi\|^2 \\ &= \sum_{t=p+1}^{n} \left(X_t - \sum_{k=1}^{p} \phi_k X_{t-k} \right)^2 \\ &= \sum_{t=p+1}^{n} \epsilon_t^2.\end{aligned}$$

- This is a standard least squares problem.

Definition 4.2 (Forward least squares estimators)

We write the minimiser by $\widehat{\phi}_F$, the estimator takes the form

$$\widehat{\phi}_F = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{X}_F \tag{4.6}$$

which is the usual least squares estimator. Finally we may estimate σ_ϵ^2 using standard least squares methods:

$$\widehat{\sigma}_F^2 = \left\|\mathbf{X}_F - \mathbf{F}\widehat{\phi}_F\right\|^2/(n-p-p). \tag{4.7}$$

Together, $\widehat{\phi}_F, \widehat{\sigma}_F^2$ are known as the forward least squares estimators.

Backward least squares: time reversibility

The backward least squares approach is based on the time reversibility property of a Gaussian AR process.

- In particular, there exists a representation $Y_t = X_{-t}$ as an AR process with the same parameters.
- Specifically,

$$Y_t = \sum_{j=1}^{p} \phi_{j,p} Y_{t-j} + \tilde{\nu}_t \tag{4.8}$$

  where $\tilde{\nu}_t$ has the same distribution as ϵ_t. Rewriting in terms of X_t, with $\nu_t = \tilde{\nu}_{-t}$ we have

$$X_t = Y_{-t} = \sum_{j=1}^{p} \phi_{j,p} Y_{-t-j} + \tilde{\nu}_{-t} = \sum_{j=1}^{p} \phi_{j,p} X_{t+j} + \nu_t.$$

The proof of (4.8) will be given later, using frequency domain methods.

Backward least squares: formulation

Now perform the same trick as in the forwards case:

$$\begin{aligned} X_1 &= \phi_1 X_2 + \phi_2 X_3 + \cdots + \phi_p X_{p+1} + \nu_1 \\ &\vdots \\ X_{n-p} &= \phi_1 X_{n-p+1} + \phi_2 X_{n-p+2} + \cdots + \phi_p X_n + \nu_n \end{aligned} \tag{4.9}$$

- This time we are regressing the first $n-p$ values in the time series against the values at appropriate lags.
- In contrast, the forwards approach used the last $n-p$ values.

Backward least squares: matrix form

In matrix form

$$\mathbf{X}_B = \mathbf{B}\boldsymbol{\phi} + \boldsymbol{\nu}$$

where

$$\mathbf{X}_B = \begin{pmatrix} X_1 \\ \vdots \\ X_{n-p} \end{pmatrix}, \quad \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}, \quad \boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_{n-p} \end{pmatrix}$$

and

$$\mathbf{B} = \begin{pmatrix} X_2 & X_3 & \cdots & X_{p+1} \\ X_3 & X_4 & \cdots & X_{p+2} \\ \vdots & \vdots & \ddots & \cdots \\ X_{n-p+1} & X_{n-p+2} & \cdots & X_n \end{pmatrix}.$$

Definition 4.3 (Backward least squares estimators)

We write the minimiser by $\widehat{\phi}_B$, the estimator takes the form

$$\widehat{\phi}_B = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{X}_B \tag{4.10}$$

which is again the least squares estimator. Finally we may estimate $\sigma_\epsilon^2 = \sigma_\nu^2$ using standard least squares methods:

$$\widehat{\sigma}_B^2 = \left\|\mathbf{X}_B - \mathbf{B}\widehat{\phi}_B\right\|^2 / (n - p - p). \tag{4.11}$$

Together, $\widehat{\phi}_F, \widehat{\sigma}_B^2$ are known as the backward least squares estimators.

- Analogously to the forward case, $\widehat{\phi}_B$ is the vector which minimises

$$\mathrm{SS}_B(\phi) = \|\mathbf{X}_B - \mathbf{B}\phi\|^2 = \sum_{t=1}^{n-p}\left(X_t - \sum_{k=1}^{p}\phi_k X_{t+k}\right)^2.$$

Forward-backward least squares

Definition 4.4 (Forward-backward least squares estimators)

The vector $\widehat{\phi}_{FB}$ that minimises

$$\mathrm{SS}_B(\phi) + \mathrm{SS}_F(\phi) \tag{4.12}$$

is called the forward-backward least squares estimator.

- Simulation studies indicate that it performs better than forward least squares or backward least squares.

Summary of least squares for AR models

- We have seen three different methods to estimate AR parameters via least squares.
- $\widehat{\phi}_{FB}, \widehat{\phi}_{B}$ and $\widehat{\phi}_{F}$ produce estimated models which need not be stationary:
  - $\rightarrow$ This may be a concern for prediction (see Lecture 10).
  - $\rightarrow$ But for other purposes, e.g. spectral estimation (see Lecture 8), the parameter values will still produce a valid estimates.

We will see stationarity conditions for AR models in Lecture 4.

Least squares for general ARMA models

Least squares for regression type models

- We now consider estimation for ARMA in general.
- We shall estimate them by minimising sums of squares of the residuals.
- Example: say we have an AR(1) model with an unknown mean

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \epsilon_t. \tag{4.13}$$

If we were to know μ and ϕ then we can work out the random perturbation by

$$\epsilon_t = Y_t - \mu - \phi(Y_{t-1} - \mu), \quad t = 2, \ldots, n. \tag{4.14}$$

Least Squares: score equations for AR(1)

We seek estimates of μ and ϕ. We thus minimise the least squares of

$$S(\mu,\phi)=\sum_{t=2}^{n}\epsilon_t^2=\sum_{t=2}^{n}\left(Y_t-\mu-\phi\left(Y_{t-1}-\mu\right)\right)^2.$$

- How do we find estimates of the parameters?
- We use calculus:

$$\frac{\partial S(\mu,\phi)}{\partial\mu}=-2(1-\phi)\sum_{t=2}^{n}\left(Y_t-\mu-\phi\left(Y_{t-1}-\mu\right)\right),$$

$$\frac{\partial S(\mu,\phi)}{\partial\phi}=-2\sum_{t=2}^{n}\left(Y_t-\mu-\phi\left(Y_{t-1}-\mu\right)\right)\left(Y_{t-1}-\mu\right).$$

We see that there is one solution to this; $\phi=1$. This is not allowed as it is non-stationary.

Least Squares: solving the score equations

Thus our estimates satisfy

$$0 = \sum_{t=2}^{n} \left(Y_t - \widehat{\mu} - \widehat{\phi}\,(Y_{t-1} - \widehat{\mu}) \right),$$

$$0 = \sum_{t=2}^{n} \left(Y_t - \widehat{\mu} - \widehat{\phi}\,(Y_{t-1} - \widehat{\mu}) \right) (Y_{t-1} - \widehat{\mu}) .$$

- We can arrive at closed form expressions:

$$\widehat{\mu} = \frac{1}{(n-1)(1-\widehat{\phi})} \left(\sum_{t=2}^{n} Y_t - \widehat{\phi} \sum_{t=2}^{n} Y_{t-1} \right)$$

$$\widehat{\phi} = \frac{\sum_{t=2}^{n} \{Y_{t-1} - \widehat{\mu}\} \{Y_t - \widehat{\mu}\}}{\sum_{t=1}^{n-1} \{Y_t - \widehat{\mu}\}^2}$$

Summary of the AR case

- For n very large, this is approximately the same as $\widehat{\mu} \approx \bar{Y}$ and $\widehat{\phi} \approx \rho_1$.
- So in this case the least-squares estimation method gives similar results as the method of moments.
- This is true for most AR(p) models.

Least Squares for MA models

Now say we want to estimate θ in an MA(1) model

$$Y_t = \epsilon_t - \theta\epsilon_{t-1}.$$

Again we minimise over θ the sum of squares

$$S(\theta) = \sum_{t=1}^{n} \left(Y_t + \theta\epsilon_{t-1}\right)^2.$$

But ϵ_{t-1} is not observed!

- Solution: assume that $\epsilon_0 = 0$ and then, for every θ we calculate estimates of $\{\epsilon_t\}$ recursively using the formula

$$\epsilon_t = Y_t + \theta\epsilon_{t-1}.$$

Least Squares for MA models: estimating innovations

- Thus

$$\begin{aligned}\hat{\epsilon}_1 &= Y_1,\\ \hat{\epsilon}_2 &= Y_2 + \theta\hat{\epsilon}_1,\\ \hat{\epsilon}_3 &= Y_3 + \theta\hat{\epsilon}_2,\\ &\vdots\\ \hat{\epsilon}_n &= Y_n + \theta\hat{\epsilon}_{n-1}.\end{aligned}$$

- In order to calculate the sum of squares $S(\theta)$ we replace the unobserved errors ϵ_t with the estimates $\hat{\epsilon}_t$.

Least Squares for ARMA models

- Consider an ARMA$(1, 1)$ model:

$$Y_t = \phi Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}.$$

- In this case we can estimate

$$S(\phi, \theta) = \sum_{t=1}^{n} \left(Y_t - \phi Y_{t-1} + \theta \varepsilon_{t-1}\right)^2.$$

- We can estimate the residuals one-by-one recursively

$$\varepsilon_t = Y_t - \phi Y_{t-1} + \theta \varepsilon_{t-1}.$$

- The recursive estimation is initiated at $Y_0 = \varepsilon_0 = 0$.

Least Squares: takeaways

- This method will work for any ARMA(p, q) model.
- For models that have an MA part, one needs to assume that q terms are zero.