Time Series lecture 10

Forecasting

Sofia Olhede

April 30, 2025

Lecture outline

1. Forecasting
2. Prediction for causal ARMA models
3. Prediction for integrated processes
4. Prediction uncertainty

Forecasting

What is forecasting

The purpose of time series analysis is often to forecast future values based on the data collected up to the present. To do this we

1. assume the form of the model;
2. estimate the parameters of the model;
3. use the assumed model structure and the estimated parameters to form predictions.

For now, we assume that the model structure and the parameters are known, and predict h lags ahead, i.e. X_{n+h} based on $X_1, \ldots, X_n$.

- $h > 0$ is called the lead or the horizon
- n is called the origin

Conditional expectation

Let $g(X_1, \ldots, X_n)$ be an estimator of X_{n+h} based on $X_1, \ldots, X_n$. We assess its performance with the prediction mean squared error

$$P_{n+h}^n := \mathbb{E}\left[(X_{n+h} - g(X_1, \ldots, X_n))^2\right]. \tag{10.1}$$

Lemma 10.1

The conditional expectation

$$g(X_1, \ldots, X_n) = \mathbb{E}\left[X_{n+h} \mid X_1, \ldots, X_n\right]$$

minimizes the prediction mean square error for X_{n+h} based on $X_1, \ldots, X_n$.

Proof of Lemma 10.1.

For Y a real-valued random variable with $\mu = \mathbb{E}[Y]$, we have

$$\begin{aligned} \mathrm{MSE}(c) &= \mathbb{E}\left[(Y-c)^2\right] \\ &= \mathbb{E}\left[(Y-\mu+\mu-c)^2\right] \\ &= \mathbb{E}\left[(Y-\mu)^2\right] + (\mu-c)^2 \\ &= \mathrm{Var}(Y) + (\mu-c)^2 . \end{aligned}$$

Thus $\mu = \arg\min_c \mathrm{MSE}(c)$. Then,

$$\begin{aligned} \mathrm{MSE}(g(X_1,\ldots,X_n)) &= \mathbb{E}\left[(X_{n+h} - g(X_1,\ldots,X_n))^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[(X_{n+h} - g(X_1,\ldots,X_n))^2 \mid X_1,\ldots,X_n\right]\right] \end{aligned}$$

will be minimised at $g(X_1,\ldots,X_n) = \mathbb{E}[X_{n+h}|X_1,\ldots,X_n]$. □

Best linear predictor

Recall the definition of linear predictor:

Definition 10.2

The best linear predictor of X_{n+h} based on $X_1, \ldots, X_n$ is the linear function

$$X_{n+h}^n = \mathcal{P}_{X_1,\ldots,X_n}(X_{n+h}) = \sum_{j=1}^{n} \beta_j X_j$$

that minimizes the prediction mean square error.

- For Gaussian time series, the conditional expectation is linear in the data.
- But in general, the conditional expectation will be a nonlinear function of the data.

For simplicity, we restrict attention to linear predictors.

Example: Gaussian AR(1)

Example 10.3

Suppose X_t is such that $X_t = \phi X_{t-1} + \varepsilon_t$, with $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
For $h > 0$, the best linear predictor of X_{n+h} based on $X_1, \ldots, X_n$ is

$$\begin{aligned} X_{n+h}^n &= \mathbb{E}\left[X_{n+h} \mid X_1, \ldots, X_n\right] \\ &= \mathbb{E}\left[\phi X_{n+h-1} + \varepsilon_{n+h} \mid X_1, \ldots, X_n\right] \\ &= \phi^h X_n \end{aligned}$$

Since X_t is stationary, $|\phi| < 1$ and $X_{n+h}^n \to 0$ as $h \to \infty$.

- X_{n+h}^n reverts to the mean of the process.
- This is a general property of stationary ARMA models.

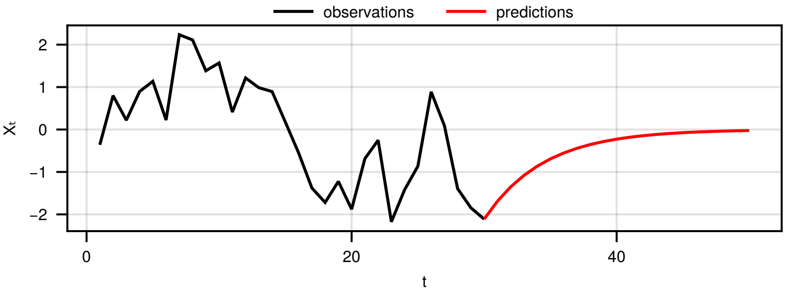


Figure: Predictions of an AR(1) process with $\phi = 0.8$ and $\sigma^2 = 1$.

Example: Gaussian MA(1)

Example 10.4

Suppose X_t is such that $X_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}$, with $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
For $h > 0$, we get

$$\begin{aligned} X_{n+h}^n &= \mu + \mathbb{E}\left[\varepsilon_{n+h} \mid X_1, \ldots, X_n\right] - \theta\mathbb{E}\left[\varepsilon_{n+h-1} \mid X_1, \ldots, X_n\right] \\ &= \mu - \theta\mathbb{E}\left[\varepsilon_{n+h-1} \mid X_1, \ldots, X_n\right], \end{aligned}$$

as ε_t are strictly uncorrelated with $X_1, \ldots, X_n$ for $t > n$.
For $h > 1$, this reduces to

$$X_{n+h}^n = \mu.$$

Prediction equations

The best linear predictor depends only on the second-order properties.

Theorem 10.5

For a zero-mean stationary process $\{X_t\}$, X_{n+h}^n is found by solving for $\beta_1, \ldots, \beta_n$ the prediction equations

$$\mathbb{E}\left[\left(X_{n+h} - X_{n+h}^n\right) X_k\right] = 0, \quad k = 1, \ldots, n.$$

The β_j are then given by the solution of the system of equations

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \gamma_{n+h-1} \\ \gamma_{n+h-2} \\ \vdots \\ \gamma_h \end{pmatrix}. \qquad (10.2)$$

- The proof for the non-zero mean processes is an exercise.

Application to ARMA processes

Rewriting the prediction equations (10.2) in matrix form

$$\mathbf{\Gamma}_n \boldsymbol{\beta} = \boldsymbol{\gamma}_{[h]},$$

we get that if $\mathbf{\Gamma}_n$ is invertible

$$X_{n+h}^n = \boldsymbol{\gamma}_{[h]}^{\mathrm{T}} \mathbf{\Gamma}_n^{-1} \boldsymbol{X} \text{ and } P_{n+h}^h = \gamma_0 - \boldsymbol{\gamma}_{[h]}^{\mathrm{T}} \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_{[h]}$$

with $\boldsymbol{X} = (X_1, \dots, X_n)^{\mathrm{T}}$.

- $\mathbf{\Gamma}_n$ is invertible for invertible ARMA models, but not in all other cases. However the best prediction X_{n+h}^n is always unique.

Comments on prediction equations

- The calculation of the solution $X_{n+h}^{n} = \gamma_{[h]}^{\mathrm{T}} \mathbf{\Gamma}_n^{-1} \boldsymbol{X}$ is inefficient when n is large because the $n \times n$ matrix $\mathbf{\Gamma}_n$ must be inverted.
- Recursive algorithms that do not require any matrix inversion have been proposed: the Durbin-Levinson and the innovations algorithms are discussed in Shumway and Stoffer (2000, §3.5).
- Theorem 10.5 is valid for any stationary process. In the next section, we focus on predictions for causal ARMA models.

Prediction for causal ARMA models

Prediction based on linear representation

We use the linear process representation and the implicit definition from an ARMA equation.

Theorem 10.6

The best linear predictor X^n_{n+h} for X_{n+h} in a causal ARMA process with general linear representation $\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ is

$$X^n_{n+h} = \sum_{j=h}^{\infty} \psi_j \varepsilon_{n+h-j} = \psi_h \varepsilon_n + \psi_{h+1} \varepsilon_{n-1} + \cdots .$$

The corresponding prediction mean square error is $\sigma^2 \sum_{j=0}^{h-1} \psi_j^2$.

- X^n_{n+h} can be evaluated from its linear representation, with unrealised innovations $\varepsilon_{n+1}, \ldots, \varepsilon_{n+h} \equiv 0$.
- The proof is an exercise for this week

Truncated predictions for ARMA models

In practice it is simpler to predict based on $\Phi(\mathrm{B})X_{t+h} = \Theta(\mathrm{B})\varepsilon_{t+h}$ than to obtain the general linear representation.
The residuals $\widehat{\varepsilon}_t$ and fitted values/predictions $\tilde{X}_t^n$ are obtained recursively

$$\widehat{\varepsilon}_t = \begin{cases} \sum_{i=1}^{p} \phi_i \tilde{X}_{t-i}^n + \left(\theta_1 \widehat{\varepsilon}_{t-1} + \cdots + \theta_q \widehat{\varepsilon}_{t-q}\right), & t = 1, \ldots, n, \\ 0, & \text{otherwise} . \end{cases}$$

$$\tilde{X}_{t+h}^n = \begin{cases} \sum_{i=1}^{p} \phi_i \tilde{X}_{t+h-i}^n - \sum_{j=1}^{q} \theta_j \widehat{\varepsilon}_{t+h-j}, & t+h > n, \\ X_{t+h}, & t+h = 1, \ldots, n, \\ 0, & t+h \leq 0. \end{cases} \tag{10.3}$$

- The forecast error is estimated by the expression in Theorem 10.6.
- Note that $\tilde{X}_t^n$ is an approximation as we had to truncate.

Comments

The best linear predictor for a causal process with mean μ has form

$$X_{n+h}^{n} = \mu + \sum_{j=h}^{\infty} \psi_j \varepsilon_{n+h-j}.$$

For a causal ARMA model $\psi_j \to 0$ exponentially fast:

$$X_{n+h} \to \mu \text{ as } h \to \infty.$$

Likewise for large h, the prediction error

$$\mathbb{E}\left[\left(X_{n+h} - X_{n+h}^{n}\right)^2\right] = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \to \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma_0, \quad h \to \infty.$$

Prediction for integrated processes

Truncated point forecasts example for ARIMA(3,1,1)

Example: ARIMA(3,1,1)

We have

$$\left(1-\phi_1 \mathrm{B}-\phi_2 \mathrm{B}^2-\phi_3 \mathrm{B}^3\right)(1-\mathrm{B}) X_t=\left(1-\theta_1 \mathrm{B}\right) \varepsilon_t,$$

which we can rewrite as

$$X_t=\underbrace{\left(1+\phi_1\right) X_{t-1}-\left(\phi_1-\phi_2\right) X_{t-2}-\left(\phi_2-\phi_3\right) X_{t-3}-\phi_3 X_{t-4}}_{\text{AR}}$$
$$+\underbrace{\varepsilon_t-\theta_1 \varepsilon_{t-1}}_{\text{MA}}.$$

We will deal with the AR and MA part as we did for ARMA models:

- AR part: replace $X_{t-1}, \ldots, X_{t-4}$ with $X_{t-1}^n, \ldots, X_{t-4}^n$.
- MA part: replace ε_t with the residuals if $t \leq n$, and otherwise set to 0.

Example: ARIMA(3,1,1) (cont.)

Let $\hat{\varepsilon}_n$ be the last observed residual (similar to eq. (10.3)). For the one-step ahead forecast we get

$$X^n_{n+1} = (1+\phi_1)X_n - (\phi_1 - \phi_2)X_{n-1} - (\phi_2 - \phi_3)X_{n-2} - \phi_3 X_{n-3} - \theta_1 \hat{\varepsilon}_n.$$

For $t > n+1$, the best linear predictor for X^n_{n+h} based on $X_1, \dots, X_n$ is then

$$X^n_t = (1+\phi_1)X^n_{t-1} - (\phi_1 - \phi_2)X^n_{t-2} - (\phi_2 - \phi_3)X^n_{t-3} - \phi_3 X^n_{t-4}.$$

- We can generalize this example to any ARIMA(p, d, q) model.

General case: falling back to ARMA

Let $\{X_t\}$ be ARIMA(p, d, q), i.e.

$$\Phi(\mathrm{B})(1-\mathrm{B})^d X_t = \Theta(\mathrm{B})\varepsilon_t.$$

Then $Z_t = \nabla^d Y_t$ is ARMA(p, q)

$$Z_t = \sum_{j=1}^{p} \phi_j Z_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

We know how to predict Z_{n+h} based on $Z_1, \ldots, Z_n$, and we can use them to predict X_{n+h} based on $X_1, \ldots, X_n$:

$$\nabla^d X_{n+h}^n = Z_{n+h}^n.$$

predicting X based on Z

Suppose that X_t ARIMA$(p, 1, q)$. Then we have $X_{n+h} - X_{n+h-1} = Z_{n+h}$ and

$$X_{n+h} = X_{n+h-1} + Z_{n+h} = \cdots = X_n + \sum_{j=0}^{h-1} Z_{n+h-j}, \quad h > 1,$$

giving

$$X_{n+h}^n = X_n + \sum_{j=0}^{h-1} Z_{n+h-j}^n, \quad h > 1.$$

If $\{Z_t\}$ has mean $\alpha \neq 0, Z_{n+h}^n \to \alpha$ as $h \to \infty$, and $X_{n+h}^n \approx X_n + \alpha h$ for h large.

- In general for ARIMA models $P_{n+h}^n \to \infty$ as $h \to \infty$.
- For SARIMA models with $d = D = 1$ and $\alpha = 0$, it can be shown that the forecast will be linear and seasonal in h.

Prediction uncertainty

Source of uncertainty

There are three components of uncertainty in predictions:

1. model uncertainty,
2. estimation uncertainty,
3. innovation uncertainty.

- We assume that the model is known and that the parameters are estimated without error. This is equivalent to removing the first two sources of uncertainty.
- There are techniques (bootstrap) to account for the model and estimation uncertainties, but we will not cover them in this course.

Innovation uncertainty for ARMA models

Let $e_n(h)$ be the h-step ahead forecast error for X_{n+h} based on $X_1, \ldots, X_n$. Then using theorem 10.6 we have

$$e_n(h) = X_{n+h} - X_{n+h}^n = \sum_{j=0}^{h-1} \psi_j \varepsilon_{n+h-j}.$$

We have $\mathbb{E}\left[e_n(h)\right] = 0$ and

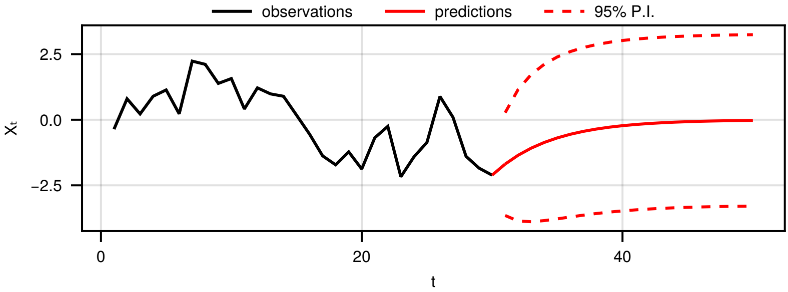
$$\mathrm{Var}\left(e_n(h)\right) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \to \gamma_0, h \to \infty.$$

Prediction intervals

If we consider a Gaussian process, we can construct a $1-\alpha$ prediction interval for X_{n+h} based on $X_1, \ldots, X_n$ as

$$X_{n+h}^{n} \pm z_{1-\alpha/2} \operatorname{Var}\left(e_n(h)\right)^{1/2},$$

where z_α is the αth percentile of the Gaussian distribution.



Bibliography

Shumway, R. H. and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3. Springer.