Financial Econometrics

Lecture 5

Fu Ouyang

April 4, 2018

Outline

Information Criteria

Model diagnostics via residual analysis tends to favor complex models as more regressors can often improve the goodness of fit (GoF). But this may give an *overfitting* model. Including irrelevant regressors may:

- inflate the SE of the estimated parameters, and/or
- make the interpretation of the fitted model difficult.

A good strategy for model selection is to find a balance between the GoF and the complexity of the model.

Akaike's information criterion (AIC) is proposed for model selection based on the trade-off between these two factors:

$$AIC = -2l(\widehat{\theta}) + 2K$$

where $l(\cdot)$ is the log-likelihood function, $\widehat{\theta}$ is the MLE, and K is the number of estimated parameters.

Information Criteria

An alternative to AIC is the *Bayesian information criterion* (BIC):

$$BIC = -2l(\widehat{\theta}) + K \log T$$

For a stationary ARMA(p, q) model with $\epsilon_t \sim N(0, \sigma^2)$,

$$AIC(p, q) = T \log(\widehat{\sigma}^2) + 2(p + q + 1)$$

$$BIC(p, q) = T \log(\widehat{\sigma}^2) + (p + q + 1) \log T$$

- The first term is a measure of the GoF.
- The second term is a penalty for the complexity of the model.

In practice, one can select the order (p, q) minimizing AIC(p, q) or BIC(p, q).

Information Criteria and Model Identification

Model Selection via AIC/BIC

- AIC tends to overestimate the orders, while BIC tends to give oversimplified models ($\log T > 2$).
- For interpretation purpose, one can use BIC to select a simple model. For forecasting purpose, one can try AIC first as a slightly more complex model causes little harm.
- AIC is asymptotically efficient, while BIC is not.
- Compute AIC and select a bunch of competitive candidates (with AIC close to to the minimum AIC).
- Select optimal model from these candidates based on interpretation, simplicity, diagnostic checking among other considerations.

Outline

Stochastic and Deterministic Trends

The following two processes are both nonstationary and exhibit similar (sample) ACF:

- Stochastic trend: $X_t = X_{t-1} + \epsilon_t$ (random walk)
- Deterministic trend: $Y_t = 0.1t + \epsilon_t$

One can use ARIMA model to handle stochastic trends, e.g.,

$$X_t = X_{t-1} + \epsilon_t + \theta\epsilon_{t-1}, |\theta| < 1 \Rightarrow \nabla X_t \sim \text{MA}(1)$$

ARIMA model can also be used to remove deterministic trends, e.g.,

$$Y_t = at + \epsilon_t \Rightarrow \nabla Y_t = a + \epsilon_t - \epsilon_{t-1} \sim \text{MA}(1)$$

$$Y_t = at^2 + bt + c + \epsilon_t \Rightarrow \nabla^2 Y_t = 2a + \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2} \sim \text{MA}(2)$$

But differencing data complicates the interpretation of the fitted model.

ADF Tests

When assuming the existence of a deterministic trend makes sense, one can also explicitly model it

$$X_t = at + c + bX_{t-1} + \epsilon_t$$

The slow decay of the sample ACF is often thought of as a symptom of nonstationarity due to the existence of either stochastic or deterministic trend.

Differencing is often used to restore stationarity and results in some ARIMA model for fitting of the data.

The *augmented Dickey-Fuller* (ADF) *test* is used to determine if differencing is necessary or identify the existence of a deterministic trend.

ADF Tests

The ADF test is based on the following models:

$$X_t = \alpha X_{t-1} + b_1 \nabla X_{t-1} + \cdots + b_p \nabla X_{t-p} + \epsilon_t$$

$$X_t = \mu + \alpha X_{t-1} + b_1 \nabla X_{t-1} + \cdots + b_p \nabla X_{t-p} + \epsilon_t$$

$$X_t = \mu + \beta t + \alpha X_{t-1} + b_1 \nabla X_{t-1} + \cdots + b_p \nabla X_{t-p} + \epsilon_t$$

The null hypothesis is $H_0 : \alpha = 1$ and the alternative hypothesis is $H_1 : \alpha < 1$. Define the testing statistic

$$W = \frac{\widehat{\alpha} - 1}{\widehat{SE}(\widehat{\alpha})}$$

The according critical values can be found in page 25 of the textbook.

Seasonality

Some financial time series are characteristic of periodic behavior. A typical example is the *seasonality*, i.e., the data exhibit seasonal pattern.

One can use the seasonal difference to remove the seasonal effect, e.g.,

$$\nabla_4 X_t \equiv (1 - B^4) X_t = X_t - X_{t-4}$$

More generally, seasonal difference with periodicity m applies the operator $\nabla_m \equiv (1 - B^m)$.

Stochastic trend, deterministic trend and seasonality can be removed simultaneously, e.g.,

$$(1 - B^4)(1 - B) X_t = (X_t - X_{t-4}) - (X_{t-1} - X_{t-5})$$

The resulting series then can be fitted directly to an ARMA model.

Outline

Information Criteria and Model Identification

Trends and Seasonality

Forecasting

Forecasting

One of the primary goals in time series analysis is to forecast the future values based on the historical data.

Consider predicting X_{T+k} for $k \geq 1$ based on $X^T \equiv \{X_T, X_{T-1}, ..., X_1\}$.

Denote by $X_T(k)$ the predictor for X_{T+k}. When $X_T(k)$ minimizes the *mean squared predictive error* (MSPE), i.e.,

$$X_T(k) = \arg \inf_{f \in \mathcal{F}} E[(X_{T+k} - f(X^T))^2]$$

$X_T(k)$ is called the *least squares predictor*. It is not hard to show that

$$X_T(k) = E[X_{T+k}|X^T]$$

when $E[X_{T+k}|X^T] = \beta_0 + \beta_1 X_1 + \cdots + \beta_T X_T$, $X_T(k)$ is called the *best linear predictor*.

Forecasting AR(1) Processes

Consider an AR(1) model:

$$X_t = bX_{t-1} + \epsilon_t$$

where $|b| < 1$ and $\epsilon_t \sim WN(0, \sigma^2)$. Assume the unpredictable condition (mean independence)

$$E[\epsilon_t | X_{t-1}, X_{t-2}, ...] = 0$$

Let $E_T[\cdot] \equiv E[\cdot|\Omega_T]$ be the conditional expectation given the information up to time T, Ω_T. Here $\Omega_T = \{X_T, X_{T-1}, ...\}$.

Then the one-step ahead predictor and its MSPE are

$$X_T(1) = E_T[bX_T + \epsilon_{T+1}] = bX_T$$

$$\text{MSPE}[X_T(1)] = E[(X_{T+1} - X_T(1))^2] = E(\epsilon_{T+1}^2) = \sigma^2$$

Forecasting AR(1) Processes

More generally, for any $k \geq 1$, by recursive substitution,

$$X_{T+k} = b^k X_T + \epsilon_{T+k} + b\epsilon_{T+k-1} + \cdots + b^{k-1}\epsilon_{T+1}$$

Then

$$X_T(k) = E_T[b^k X_T + \epsilon_{T+k} + b\epsilon_{T+k-1} + \cdots + b^{k-1}\epsilon_{T+1}] = b^k X_T$$

and

$$\begin{aligned} \text{MSPE}[X_T(k)] &= E[(X_{T+k} - X_T(k))^2] \\ &= E[(\epsilon_{T+k} + b\epsilon_{T+k-1} + \cdots + b^{k-1}\epsilon_{T+1})^2] \\ &= (1 + b^2 + \cdots + b^{2(k-1)})\sigma^2 \end{aligned}$$

Hence

$$X_{T+k} \overset{a}{\sim} N(X_T(k), (1 + b^2 + \cdots + b^{2(k-1)})\sigma^2)$$

Forecasting AR(1) Processes

Summary for AR(1) Forecasting

$$X_T(k) = b^k X_T$$

$$\text{MSPE}[X_T(k)] = (1 + b^2 + \cdots + b^{2(k-1)})\sigma^2$$

$$X_{T+k} \overset{a}{\sim} N(X_T(k), (1 + b^2 + \cdots + b^{2(k-1)})\sigma^2)$$

When k increases (long term forecasting),

- $\text{MSPE}[X_T(k)] \nearrow$.
- $X_T(k) \xrightarrow{p} 0 = E(X_{T+k})$, as $k \to \infty$.
- $\text{MSPE}[X_T(k)] \to \sigma^2/(1 - b^2) = Var(X_{T+k})$, as $k \to \infty$.

In practice, both b and σ^2 are replaced by their estimates.

Forecasting ARMA(p, q) Processes

Consider a causal and invertible ARMA(p, q) model

$$X_t - b_1 X_{t-1} - \cdots - b_p X_{t-p} = \epsilon_t + a_1 \epsilon_{t-1} + \cdots + a_q \epsilon_{t-q}$$

where $\epsilon_t \sim WN(0, \sigma^2)$ and $E(\epsilon_t | X_{t-1}, X_{t-2}, ...) = 0$.

The invertible condition implies that X_t can be represented as an AR(∞) process:

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \varphi_j X_{t-j}$$

where all φ_j are determined (exclusively) by $(a_1, ..., a_q, b_1, ..., b_p)$. Hence, ϵ^t can be recovered by X^t for all t and

$$\epsilon_T(k) \equiv E_T(\epsilon_{T+k}) = 0 \cdot \mathbf{1}[k \geq 1] + \epsilon_{T+k} \cdot \mathbf{1}[k \leq 0]$$

Forecasting ARMA(p, q) Processes

Recall that

$$X_{T+1} = b_1 X_T + \cdots + b_p X_{T+1-p} + \epsilon_{T+1} + a_1 \epsilon_T + \cdots + a_q \epsilon_{T+1-q}$$

then the one-step ahead predictor is

$$\begin{aligned} X_T(1) &\equiv E_T(X_{T+1}) \\ &= b_1 X_T + \cdots + b_p X_{T+1-p} + a_1 \epsilon_T + \cdots + a_q \epsilon_{T+1-q} \end{aligned}$$

whose MSPE is

$$\text{MSPE}[X_T(1)] = E[(X_{T+1} - X_T(1))^2] = E(\epsilon_{T+1}^2) = \sigma^2$$

In practice, assume $X_t = 0$ for all $t \leq 0$, and then recover ϵ^T using X^T.

Forecasting ARMA(p, q) Processes

More generally, for all $k \geq 1$,

$$\begin{aligned}&X_{T+k}\\=&b_1 X_{T+k-1} + \cdots + b_p X_{T+k-p} + \epsilon_{T+k} + a_1 \epsilon_{T+k-1} + \cdots + a_q \epsilon_{T+k-q}\end{aligned}$$

$$\begin{aligned}&X_T(k) \equiv E_T(X_{T+k})\\=&b_1 X_T(k-1) + \cdots + b_p X_T(k-p) + a_1 \epsilon_T(k-1) + \cdots + a_q \epsilon_T(k-q)\end{aligned}$$

To compute MSPE$[X_T(k)]$, one can appeal to the causality assumption, i.e., X_t admits an MA(∞) representation:

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \psi_j \epsilon_{t-j}$$

where ψ_j are determined by $(a_1, ..., a_q, b_1, ..., b_p)$.

Forecasting ARMA(p, q) Processes

Then

$$X_{T+k} = \epsilon_{T+k} + \sum_{j=1}^{\infty} \psi_j \epsilon_{T+k-j}$$

$$X_T(k) = \sum_{j=k}^{\infty} \psi_j \epsilon_{T+k-j}$$

Hence

$$\text{MSPE}[X_T(k)] = E[(X_{T+k} - X_T(k))^2] = \sigma^2 \left(1 + \sum_{j=1}^{k-1} \psi_j^2\right)$$

Forecasting ARMA(p, q) Processes

Summary for ARMA(p, q) Forecasting

$$X_T(k) = \sum_{j=k}^{\infty} \psi_j \epsilon_{T+k-j}$$

$$\text{MSPE}[X_T(k)] = \sigma^2 \left(1 + \sum_{j=1}^{k-1} \psi_j^2\right)$$

When k increases (long term forecasting),

- $\text{MSPE}[X_T(k)] \nearrow$.
- $X_T(k) \xrightarrow{p} 0 = E(X_{T+k})$, as $k \to \infty$.
- $\text{MSPE}[X_T(k)] \to Var(X_{T+k})$, as $k \to \infty$.