Financial Econometrics

Lecture 4

Fu Ouyang

March 28, 2018

Outline

Random Walks

A random walk (without drift) is defined as

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

where $\epsilon_t \sim WN(0, \sigma^2)$. Using backshift operator, we have

$$(1 - B)\, X_t = \epsilon_t \Leftrightarrow X_t = \sum_{j=0}^{\infty} \epsilon_{t-j}$$

It is easy to obtain that (assume $\epsilon_t = 0$ for all $t \leq 0$)

$$Var(X_t) = \sum_{j=0}^{t-1} Var(\epsilon_{t-j}) = t\sigma^2, Cov(X_t, X_{t+k}) = Var(X_t) = t\sigma^2$$

$$Corr(X_t, X_{t+k}) = \frac{t\sigma^2}{\sqrt{t\sigma^2\,(t+k)\,\sigma^2}} = \frac{t}{\sqrt{t\,(t+k)}} \to 1, t \to \infty$$

ARIMA Models

Let ∇ denote the *difference operator*, i.e.,

$$\nabla X_t \equiv X_t - X_{t-1}$$

and

$$\nabla^d X_t \equiv \nabla(\nabla^{d-1} X_t)$$

for all integer $d \geq 1$.

If $\nabla^d X_t$ is a stationary ARMA(p, q), X_t is called an *autoregressive integrated moving average* (ARIMA) model denoted as $X_t \sim$ ARIMA(p, d, q).

Example: ARIMA(0,1,1)

$$X_t - X_{t-1} = \epsilon_t - \theta\epsilon_{t-1}, |\theta| < 1$$

where $\epsilon_t \sim WN(0, \sigma^2)$.

ARIMA Models

Example: ARIMA(0,1,1)

Using the backshift operator, it is easy to verify that $X_t \sim$AR(∞) with coefficients $(1-\theta)\theta^k$, i.e.,

$$X_t = (1-\theta)(X_{t-1} + \theta X_{t-2} + \theta^2 X_{t-3} + \cdots) + \epsilon_t$$

Note that $\sum_{k=0}^{\infty}(1-\theta)\theta^k = 1$. Hence the weights decay and sum up to 1.

The "best" predictor (in the mean square error sense) of X_{t+1} is

$$E(X_{t+1}|X_t, X_{t-1}, \ldots) = (1-\theta)(X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \cdots)$$

which is called *exponential smoothing* and widely used in practice.

See time series simulated from random walk and ARIMA(0,1,1) [R code].

Outline

Least Square Estimation for AR(p) Models

Consider the AR(p) model

$$X_t = b_0 + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \epsilon_t, \epsilon_t \sim WN(0, \sigma^2)$$

With observations $\{X_t\}_{t=1}^{T}$, we can estimate parameters $b \equiv (b_0, b_1, \ldots, b_p)$ via a linear regression:

$$\widehat{b} \equiv (\widehat{b}_0, \widehat{b}_1, ..., \widehat{b}_p) = \arg\min_{b \in \mathbb{R}^{p+1}} \sum_{t=p+1}^{T} (X_t - b_0 - b_1 X_{t-1} - \cdots - b_p X_{t-p})^2$$

which is called the *least square estimator* (LSE) for b.

Note that both $\widehat{b}$ and $\widehat{Var}(\widehat{b})$ have explicit expressions. Hypothesis tests can be conducted easily.

Least Square Estimation for AR(p) Models

Once $\widehat{b}$ is obtained, we can compute the LSE for σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{T-2p-1} \sum_{t=p+1}^{T} \left(X_t - \widehat{b}_0 - \widehat{b}_1 X_{t-1} - \cdots - \widehat{b}_p X_{t-p} \right)^2$$

where the divider is $T-2p-1$ because the effective sample size is $T-p$ and the number of parameters is $p+1$.

Least Square Estimation for ARMA(p, q) Model

Consider the ARMA(p, q) model

$$X_t = b_0 + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \epsilon_t + a_1 \epsilon_{t-1} + \cdots + a_q \epsilon_{t-q}$$

where $\epsilon_t \sim WN(0, \sigma^2)$ and ϵ_{p+1-k} is assumed to be 0 for all $1 \leq k \leq q$.

Let $a \equiv (a_1, ..., a_q)$ and $b \equiv (b_0, b_1, \ldots, b_p)$. We can compute the LSE for (a, b) using the iterative algorithm below:

Iterative Linear Approximation

(1) Start from initial values of $\epsilon_{p+1-q} = 0, ..., \epsilon_p = 0$. For $t \geq p+1$, Define

$$\epsilon_t(a, b) = X_t - b_0 - \sum_{j=1}^{p} b_j X_{t-j} - \sum_{l=1}^{q} a_l \epsilon_{t-l}(a, b)$$

Least Square Estimation for ARMA(p, q) Model

Iterative Linear Approximation

(2) Compute the following iterative estimator with some starting values $(a, b) = (\bar{a}, \bar{b})$:

$$\begin{aligned} & (\widehat{a}_k, \widehat{b}_k) \\ = \quad & \arg\min_{a,b} \sum_{t=p+1}^{T} [\epsilon_t(\widehat{a}_{k-1}, \widehat{b}_{k-1})]^2 \\ = \quad & \arg\min_{a,b} \sum_{t=p+1}^{T} [X_t - b_0 - \sum_{j=1}^{p} b_j X_{t-j} - \sum_{l=1}^{q} a_l \epsilon_{t-l}(\widehat{a}_{k-1}, \widehat{b}_{k-1})]^2 \end{aligned}$$

for $k = 1, 2, ...$, where $\epsilon_t(\cdot)$ is defined in (1).

(3) Repeat (2) till $(\widehat{a}_k, \widehat{b}_k) \approx (\widehat{a}_{k-1}, \widehat{b}_{k-1})$.

Gaussian Maximum Likelihood Estimation

If we assume $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, then $\theta \equiv (a, b, \sigma^2)$ can be more efficiently estimated using the Gaussian maximum likelihood estimation, and the resulting estimator is called the *maximum likelihood estimator* (MLE).

Consider a general ARMA(p, q) model:

$$X_t = b_0 + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \epsilon_t + a_1 \epsilon_{t-1} + \cdots + a_q \epsilon_{t-q}$$

Let $X^t \equiv (X_1, ..., X_t)$ for all $1 \leq t \leq T$. We define the likelihood function for the model as

$$\begin{aligned} L(\theta) &\equiv f(X^T; \theta) \\ &= f(X_T; \theta | X^{T-1}) \times f(X_{T-1}; \theta | X^{T-2}) \times \cdots \times f(X_{p+1}; \theta | X^p) \times f(X^p) \end{aligned}$$

Taking log and dropping the "constant" term $\log f(X^p)$ leads to the log-likelihood function for the model.

Gaussian Maximum Likelihood Estimation

$$l(\theta) \equiv \log L(\theta) - \log f(X^p) = \sum_{t=p+1}^{T} \log f(X_t; \theta | X^{t-1})$$

Given $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, $f(X_t; \theta | X^{t-1})$ has explicit expression. Then we can employ a Newton-Raphson algorithm to compute the Gaussian MLE via the following optimization procedure:

$$\widehat{\theta} = \arg\max_{\theta} l(\theta) = \arg\max_{\theta} \sum_{t=p+1}^{T} \log f(X_t; \theta | X^{t-1})$$

The covariance matrix of $\widehat{\theta}$ can be consistently estimated as

$$\widehat{Var}(\widehat{\theta}) = -\ddot{l}(\widehat{\theta})^{-1}$$

where $\ddot{l}(\cdot)$ is the Hessian matrix of the log-likelihood function $l(\cdot)$.

Gaussian Maximum Likelihood Estimation

Under mild conditions,

- Gaussian MLE is consistent and asymptotically normal.
- Gaussian MLE is often used when ϵ_t is not normal. The resulting estimator is called *quasi-MLE*.
- Gaussian MLE is more efficient than LSE as it makes (and make uses of) stronger assumption.
- Statistical inference on $g(\widehat{\theta})$ can be done using the *Delta method* when $g(\cdot)$ is differentiable.

$$\widehat{Var}(g(\widehat{\theta})) \approx -\dot{g}(\widehat{\theta})'\ddot{l}(\widehat{\theta})^{-1}\dot{g}(\widehat{\theta})$$

  where $\dot{g}(\cdot)$ is the gradient or Jacobian of $g(\cdot)$.

Now let's apply LSE and MLE to an empirical example [R code].

Outline

Model Diagnostics

Model Diagnostics is important as it can help check if the specification of a time series model is appropriate.

The key idea is straightforward: if a ARMA(p, q) model is adequate for an observed data, the residuals

$$\widehat{\epsilon}_t = X_t - \sum_{j=1}^{p} \widehat{b}_j X_{t-j} - \sum_{l=1}^{q} \widehat{a}_l \widehat{\epsilon}_{t-l}, t = p+1, ..., T$$

with $\widehat{\epsilon}_{p+1-q} = 0, ..., \widehat{\epsilon}_p = 0$ should behave like white noise.

Residual Analysis

Visual diagnostics

- Residual plot, i.e., plotting $\widehat{\epsilon}_t$ (more often standardized, $\widehat{\epsilon}_t / \widehat{SE}(\widehat{\epsilon}_t)$) against time t.

Model Diagnostics

Residual Analysis

- Plot $\widehat{\epsilon}_t$ against the fitted values $\widehat{X}_t$.
- ACF, PACF, EACF.
- QQ plots for diagnosing normality assumption.

Statistical tests for white noise

- Perform the Ljung-Box tests for white noise.

The tsdiag() function coming with R is designed for conducting time series diagnostics [R code].