Financial Econometrics

Lecture 2

Fu Ouyang

School of Finance
Nankai University

March 7, 2018

Outline

Efficient Markets Hypothesis and Statistical Models for Returns

Linear Time Series Models: Stationarity

Stationary ARMA Models: Moving Average Processes

Efficient Market Hypothesis (EMH)

Strong EMH: Security prices P_t of traded assets reflect instantly all available information up to time t, public or private. Individuals do not have comparative advantages in the acquisition of information, and so there is no arbitrage opportunities.

Semi-strong EMH: Security prices merely reflect efficiently all past public information, leaving rooms for the value of private information.

Under the EMH, an asset return process may be expressed as

$$r_t = \mu_t + \epsilon_t, E(\epsilon_t) = 0, Var(\epsilon_t) = \sigma_t^2$$

- $E(\epsilon_t) = 0 \Rightarrow \mu_t = E(r_t)$, μ_t is the *rational expectation* of r_t at time $t-1$.
- ϵ_t is called an *innovation* representing the return due to unpredictable "news" that arrives between time $t-1$ and t.

Efficient Market Hypothesis

Combining the EMH and the stationarity feature discussed in last lecture, it makes sense to assume

$$\mu_t = \mu$$

For $\{\epsilon_t\}$, there are three different types of assumptions:

1. White Noise (WN) Innovations

$\{\epsilon_t\}$ are white noise, denoted as $\epsilon_t \sim WN(0, \sigma^2)$. Under WN assumption, for all $t \neq s$,

$$Cov(\epsilon_t, \epsilon_s) = 0$$

2. Martingale Difference (MD) Innovations

For any t,

$$E(\epsilon_t | r_{t-1}, r_{t-2}, ...) = E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, ...) = 0$$

Efficient Market Hypothesis

3. IID Innovations

$\{\epsilon_t\}$ are independent and identically distributed (IID), denoted as $\epsilon_t \sim IID(0, \sigma^2)$.

IID $\Rightarrow$ MD: $E(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, ...) = E(\epsilon_t) = 0$

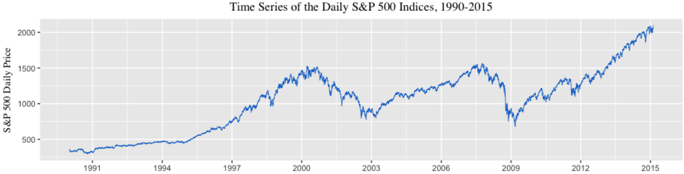
MD $\Rightarrow$ WN: For any $t > s$, by the law of iterated expectations (LIE),

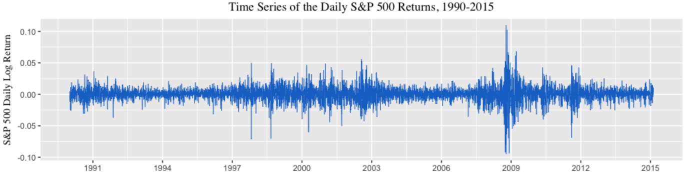
$$\begin{aligned} Cov(\epsilon_t, \epsilon_s) &= E(\epsilon_t \epsilon_s) \\ &= E[E(\epsilon_t \epsilon_s|\epsilon_{t-1}, \epsilon_{t-2}, ...)] \\ &= E[\epsilon_s E(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, ...)] \\ &= 0 \end{aligned}$$

To sum up, IID $\Rightarrow$ MD $\Rightarrow$ WN.

Revisiting S&P 500

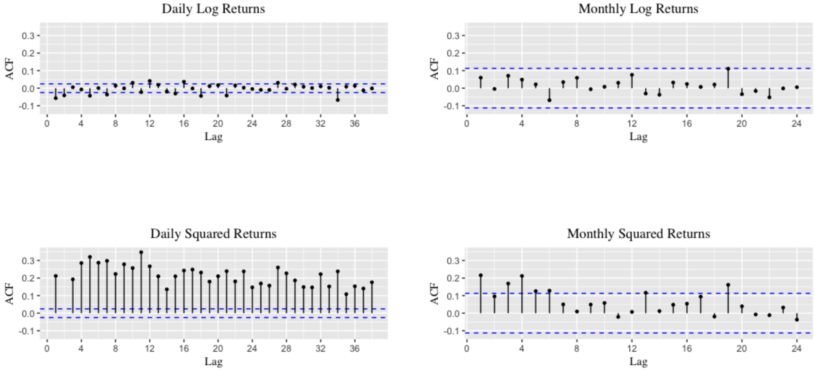
Figure 1: Time Series of P_t and r_t





Revisiting S&P 500

Figure 2: ACF of r_t and r_t^2



IID Innovations and Random Walk

Recall $r_t = \log(P_t/P_{t-1})$. Under IID assumption, $\{\log P_t\}$ form a *random walk*, i.e.,

$$\log P_t = \mu + \log P_{t-1} + \epsilon_t$$

- $\{P_t\}$ form a geometric random walk.
- If $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, $\{P_t\}$ form a log normal geometric random walk.
- As $\Delta t \to 0$, $\log P_t \rightsquigarrow$ Brownian motion ($P_t \rightsquigarrow$ geometric Brownian motion).

IID assumption is too strong to be true for modeling $\{r_t\}$ as it implies $Cov[f(r_t), f(r_s)] = 0$ for any function f (see Figure 2).

White Noise Innovations

The WN assumption is consistent with the stylized features of returns (see Figure 1 & 2) and is essential for the EMH.

Suppose $Corr(\epsilon_{t+1}, \epsilon_t) = \rho \neq 0$. Let $\widehat{r}_{t+1} \equiv \mu$ which is the fair predictor for r_{t+1} at time t under the EMH. Consider an alternative predictor

$$\tilde{r}_{t+1} = \mu + \rho(r_t - \mu)$$

We have

$$E(\tilde{r}_{t+1}) = E(\widehat{r}_{t+1}) = \mu$$

$$E[(\tilde{r}_{t+1} - r_{t+1})^2] = (1-\rho^2)\sigma^2 < \sigma^2 = E[(\widehat{r}_{t+1} - r_{t+1})^2]$$

$\tilde{r}_{t+1}$ dominates $\widehat{r}_{t+1}$ as it is the same unbiased but has smaller mean squared predictive error, which violates the EMH.

Martingale Difference Innovations

The MD assumption is consistent with the EMH. It is not hard to show that

$$E(r_{t+1}|r_t, r_{t-1}, ...) = \arg\inf_{g \in \mathcal{G}} E[(r_{t+1} - g(r_t, r_{t-1}, ...))^2]$$

and under the MD assumption

$$E(r_{t+1}|r_t, r_{t-1}, ...) = \mu + E(\epsilon_{t+1}|r_t, r_{t-1}, ...) = \mu$$

i.e., given $r_t, r_{t-1}, ...$, the best point predictor for r_{t+1} is μ.

The MD assumption is the most appropriate mathematical form of the EMH as it is assures that asset returns cannot be predicted by any rules, but allows certain nonlinear dependence.

Outline

Efficient Markets Hypothesis and Statistical Models for Returns

Linear Time Series Models: Stationarity

Stationary ARMA Models: Moving Average Processes

Linear Time Series Models

Data obtained from observations collected sequentially over time are called *time series*. The purpose of analyzing time series data:

1. Recover the *data generating process* (DGP) that generates the data.
2. Forecast the future values of a time series using historical data.

In the following couple of lectures, we will study a class of models which depict the linear features (the first two moments and linear dependence) of time series.

In what follows, we use $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ (or for notation simplicity $\{X_t\}$) to represent a generic stochastic process (i.e., a sequence of random variables). But sometimes it is convenient to refer to $\{X_t\}$ itself as a set of observed time series data.

Weak Stationarity

The assumption of stationarity plays a central role in forecasting, which in general refers to certain time invariance properties of the underlying DGP.

Weak Stationarity

$\{X_t\}$ is *weakly stationary* (or *second order stationary* or *covariance stationary*) if $E(X_t^2) < \infty$ and both $E(X_t)$ and $Cov(X_t, X_{t+k})$, for any integer k, do not depend on t.

- $E(X_t)$ is a constant, i.e., $E(X_t) = \mu$.
- $Cov(X_t, X_{t+k})$ is independent of t for all $k = 0, \pm 1, \pm 2, \cdots$.
- $|Cov(X_t, X_{t+k})| < \infty$ by $|E(X_t X_{t+k})|^2 \leq E(X_t^2)E(X_{t+k}^2)$ (recall the Cauchy-Schwarz inequality) and $E(X_t^2) < \infty$.
- $\{X_t\}$ is weakly stationary $\Leftrightarrow$ $\{X_t\}$ has finite and time-invariant first two moments.

Autocovariance Function

The *autocovariance function* (ACVF) is defined as

$$\gamma(k) = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

for $k = 0, \pm 1, \pm 2, \cdots$. Note that $\gamma(0) = Var(X_t)$ and $\gamma(k) = \gamma(-k)$.

The variance-covariance matrix of the vector $(X_t, ..., X_{t+k})$ is

$$Var(X_t, ..., X_{t+k}) = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(k) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \gamma(k-3) & \cdots & \gamma(1) \\ \gamma(k) & \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{pmatrix}$$

Autocorrelation Function

The *autocorrelation function* (ACF) is defined as

$$\rho(k) = Corr(X_t, X_{t+k}) = \gamma(k)/\gamma(0)$$

for $k = 0, \pm 1, \pm 2, \cdots$. Note that $\rho(0) = 1$ and $\rho(k) = \rho(-k)$.

Sample ACVF and sample ACF

How to use an observed sample $X_1, ...X_T$ to estimate ACVF and ACF?

$$\widehat{\gamma}(k) = \frac{1}{T}\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X}), \widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0)$$

where $\bar{X} = \sum_{t=1}^{T} X_t/T$. $\widehat{\gamma}(k)$ and $\widehat{\rho}(k)$ are called sample ACVF and sample ACF, respectively.

Note that the estimator $\widehat{\gamma}(k)$ use divisor T instead of $T - k$!

Sample ACVF and Sample ACF

Let $Z_t \equiv X_t - \bar{X}$.

$$\mathbf{Z} = \begin{pmatrix} 0 & 0 & \cdots & 0 & Z_1 & Z_2 & \cdots & Z_{T-1} & Z_T \\ 0 & 0 & \cdots & Z_1 & Z_2 & Z_3 & \cdots & Z_T & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Z_1 & Z_2 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{pmatrix}_{(k+1)\times(k+T)}$$

$$\widehat{Var}(X_t, ..., X_{t+k}) = \frac{1}{T}\mathbf{Z}\mathbf{Z}'$$

Using divisor T ensures that $\widehat{Var}(X_t, ..., X_{t+k})$ is *semi-positive definite*, i.e., for any $(k+1)$-vector v, $v'\widehat{Var}(X_t, ..., X_{t+k})v \geq 0$.

Strong Stationarity

Strong Stationarity

$\{X_t\}$ is said to be *strongly stationary* or *strictly stationary* if the joint distribution of $(X_1, ...X_k)$ is the same as that of $(X_{t+1}, ..., X_{t+k})$ for any $k \geq 1$ and t.

Note that

- Provided $E(X_t^2) < \infty$, strong stationarity $\Rightarrow$ weak stationarity.
- The strong stationarity of $\{X_t\}$ $\Rightarrow$ the strong stationarity of $\{g(X_t)\}$ for any function g.
- The assumption of strong stationarity will be needed in the context of nonlinear prediction.

Outline

Moving Average (MA) Processes: Definition

Let $\epsilon_t \sim WN(0, \sigma^2)$. For a fixed integer $q \geq 1$, we say $X_t \sim \text{MA}(q)$ if X_t is defined as a moving average of q successive ϵ_t as follows

$$X_t = \mu + \epsilon_t + \sum_{k=1}^{q} a_k \epsilon_{t-k}$$

where $\mu, a_1, ..., a_q$ are constant coefficients.

- μ is the stationary expectation of X_t, $E(X_t) = \mu$.
- $\{\epsilon_t\}$ stands for a sequence of innovations (shocks) to the market in each period.
- $\{a_k\}$ can be thought of as "discount" factors associated with lagged innovations $\{\epsilon_{t-k}\}$.
- All MA(q) processes are (weakly) stationary. (why?)

MA(q) Processes: ACVF and ACF

Recall $\rho(k) = Cov(X_{t+k}, X_t)/Var(X_t) = \gamma(k)/\gamma(0)$. Letting $a_0 \equiv 1$,

$$\gamma(0) = Var(X_t) = E\left[\left(\sum_{l=0}^{q} a_k \epsilon_{t-l}\right)^2\right]$$

$$\gamma(k) = Cov(X_{t+k}, X_t) = E\left[\left(\sum_{l=0}^{q} a_k \epsilon_{t-l}\right)\left(\sum_{l=0}^{q} a_k \epsilon_{t+k-l}\right)\right]$$

By $\epsilon_t \sim WN(0, \sigma^2)$, $E(\epsilon_t \epsilon_s) \neq 0$ if and only if $t = s$. Hence,

$$\gamma(0) = \sigma^2 \sum_{l=0}^{q} a_l^2$$

and $\forall k > q$, $Cov(X_{t+k}, X_t) = 0$, i.e., the ACF of MA(q) process cuts off at q.

MA(q) Processes: ACVF and ACF

For $1 \leq k \leq q$, common WN terms are $\epsilon_{t+k-q}, \ldots, \epsilon_{t-1}, \epsilon_t$, and so

$$\gamma(k) = \sigma^2 (a_q a_{q-k} + \ldots + a_{k+1} a_1 + a_k a_0)$$

To sum up, we have

$$\rho(k) = \frac{a_q a_{q-|k|} + \ldots + a_{|k|+1} a_1 + a_{|k|} a_0}{a_0^2 + a_1^2 + \ldots + a_q^2} \cdot \mathbb{I}[1 \leq |k| \leq q]$$

where $\mathbb{I}[\cdot]$ is an indicator function and the $|\cdot|$ is used because of the symmetry of $\rho(k)$, i.e., $Cov(X_{t+k}, X_t) = Cov(X_{t-k}, X_t)$.

MA(∞) Processes

If we permit the order q of an MA(q) process to increase to infinity, i.e.,

$$X_t = \mu + \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$$

with $\epsilon_t \sim WN(0, \sigma^2)$, we obtain a MA($\infty$) process. MA($\infty$) is well-defined (i.e., $\sum_{j=0}^{\infty} a_j \epsilon_{t-j}$ converges in mean-square) if $\sum_{j=1}^{\infty} a_j^2 < \infty$ as

$$E\left[|\sum_{j=0}^{n} a_j \epsilon_{t-j} - \sum_{j=0}^{m} a_j \epsilon_{t-j}|^2 \right] = \sum_{j=m}^{n} \sigma^2 a_j^2 \to 0$$

Using the same derivation for MA(q), we obtain that for a MA(∞) process,

$$\gamma(0) = \sigma^2 \sum_{j=0}^{\infty} a_j^2 < \infty, \gamma(k) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+|k|} < \infty$$