Financial Econometrics I

Lecture 4

Fu Ouyang

December 7, 2018

Outline

Nonstationary ARMA Processes

Forecasting

Stochastic and Deterministic Trends

Consider the following two processes:

- Stochastic trend: $X_t = X_{t-1} + \epsilon_t$ (random walk) with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
- Deterministic trend: $Y_t = 0.1t + \epsilon_t$ with $e_t \overset{iid}{\sim} N(0,1)$.

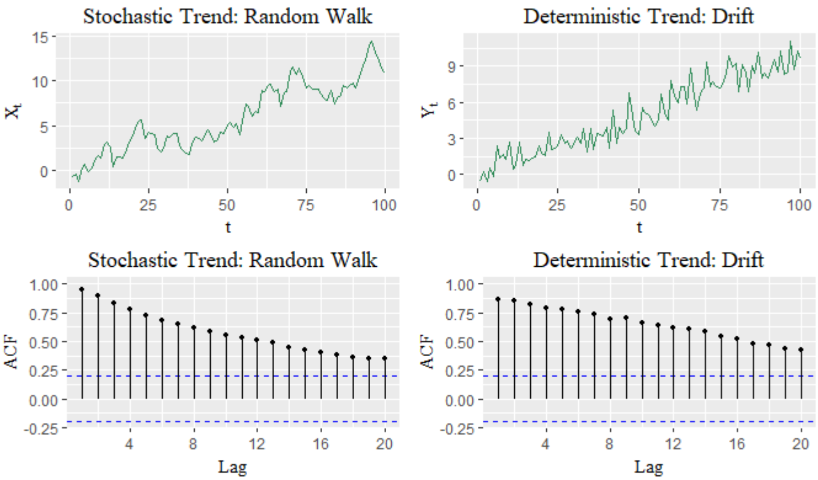
The two processes are both nonstationary as

- $Corr(X_t, X_{t+k}) = t/\sqrt{t(t+k)} > 0$.
- $E(Y_t) = 0.1t$.

It is obvious that Y_t has a deterministic trend $0.1t$. As for X_t, note that $X_t = X_0 + \sum_{s=1}^{t} \epsilon_t$, and so it has a stochastic trend $\sum_{s=1}^{t} \epsilon_t$. However, these two processes exhibit similar time series plots and ACF.

Stochastic and Deterministic Trends

Figure 1: Stochastic Trend vs. Deterministic Trend



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

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

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

Trend Removal

The slow decay of the sample ACF is often a signal for nonstationarity of a process due to either stochastic or deterministic trend.

For a process with only deterministic trend, e.g.,

$$Y_t = \beta_0 + \beta_1 t + e_t$$

One can run a linear regression and obtain the OLS estimate $\widehat{\beta}_1$ of $\beta_1$first, and then work on $Z_t = Y_t - \widehat{\beta}_1 t$.

A more commonly used (and safer) method is differencing, i.e., working on ARIMA models. For example,

- Stochastic: $X_t = X_{t-1} + \epsilon_t + \theta\epsilon_{t-1} \Rightarrow \nabla X_t = \epsilon_t + \theta\epsilon_{t-1} \sim \mathrm{MA}(1)$.
- Deterministic: $Y_t = \beta t + e_t \Rightarrow \nabla Y_t = \beta + e_t - e_{t-1} \sim \mathrm{MA}(1)$.
- Quadratic deterministic trend: $Y_t = \beta_1 t^2 + \beta_2 t + e_t \Rightarrow \nabla^2 Y_t \sim \mathrm{MA}(2)$.

ADF Tests

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 and/or identify the existence of a deterministic trend.

The ADF test is based on the following models:

- (Type 1) $X_t = \alpha X_{t-1} + \beta_1 \nabla X_{t-1} + \cdots + \beta_p \nabla X_{t-p} + \epsilon_t$.
- (Type 2) $X_t = \mu + \alpha X_{t-1} + \beta_1 \nabla X_{t-1} + \cdots + \beta_p \nabla X_{t-p} + \epsilon_t$.
- (Type 3) $X_t = \mu + \beta_0 t + \alpha X_{t-1} + \beta_1 \nabla X_{t-1} + \cdots + \beta_p \nabla X_{t-p} + \epsilon_t$.

The null hypothesis is $H_0 : \alpha = 1$ (vs. $H_1 : \alpha < 1$). Define the testing statistic

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

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

ADF Tests

Random Walk						
	Type 1		Type 2		Type 3	
	ADF	p-value	ADF	p-value	ADF	p-value
$p = 0$	0.301	0.729	-1.460	0.536	-3.190	0.094
$p = 1$	0.280	0.723	-1.460	0.536	-3.320	0.071
$p = 2$	0.305	0.730	-1.630	0.472	-3.440	0.052

Deterministic Trend						
	Type 1		Type 2		Type 3	
	ADF	p-value	ADF	p-value	ADF	p-value
$p = 0$	-0.645	0.448	-2.470	0.147	-10.730	< 0.01
$p = 1$	0.482	0.781	-1.300	0.591	-6.770	< 0.01
$p = 2$	1.135	0.930	-1.080	0.668	-5.190	< 0.01

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 $\nabla_m \equiv (1 - B^m)$.

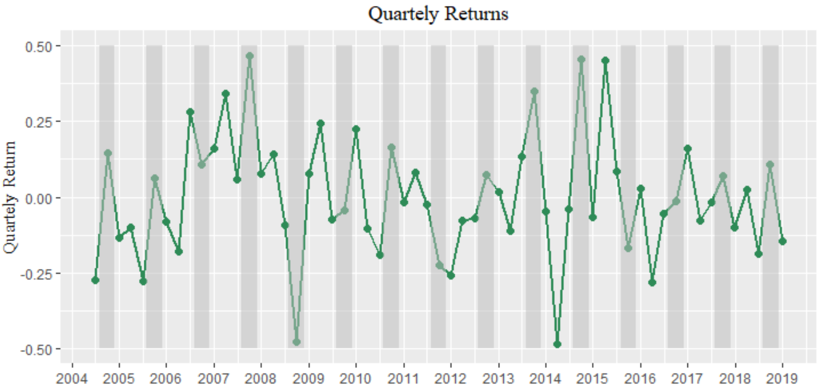
Stochastic trend, deterministic trend and seasonality can be removed simultaneously, e.g.,

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

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

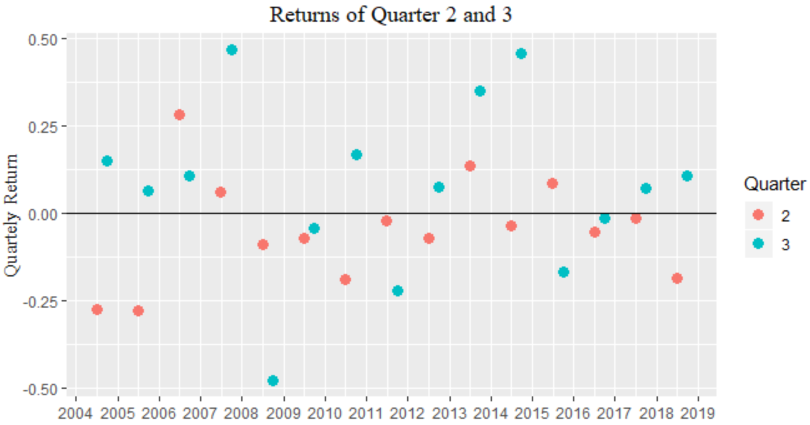
Seasonality

Figure 2: Periodic Behavior (Seasonality)



Seasonality

Figure 3: Periodic Behavior (Seasonality)



Outline

Nonstationary ARMA Processes

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

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

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

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

When k increases (long term forecasting),

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

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

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

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 $k \to \infty$.

Example: Forecasting Daily Gold Prices

Figure 4: Forecasting ARMA Models