Collective Dynamics of Financial Markets: Time series and Networks

Lecture slides

Anindya S. Chakrabarti

Economics area
Indian Institute of Management Ahmedabad
email id: anindyac@iima.ac.in

December 7, 2024

Outline

Some details about the sequence:

1. Introduction to time series
2. Modelling autocorrelation. ARMA processes
3. Stationarity, Vector AR, Granger Causality
4. Nonstationarity, Unit roots, random walk
5. Volatility clustering, Introduction to ARCH-GARCH models
6. Financial networks
7. Spectral analysis

Some examples:

Let us consider the following:

- Company sales.
- Stock market analysis.
- Inventory studies.
- Inflation forecasting.
- Weather prediction.
- Population growth.
- $\vdots$

Temperature time series

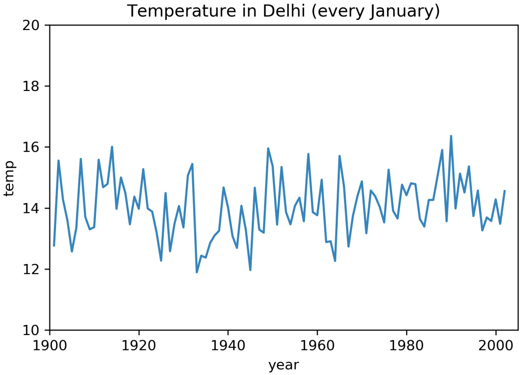


Figure: Temperature time series in Delhi over the last century in every January.

Rainfall time series

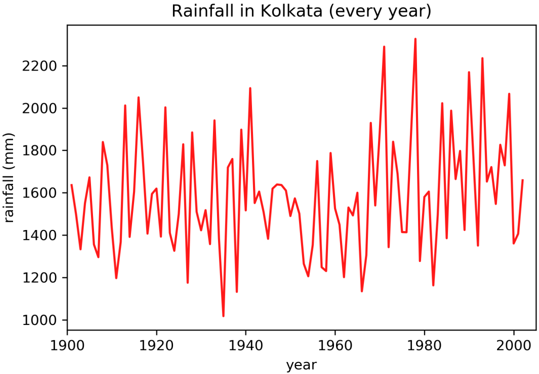


Figure: Rainfall in Kolkata over the last century.

GDP growth rate time series

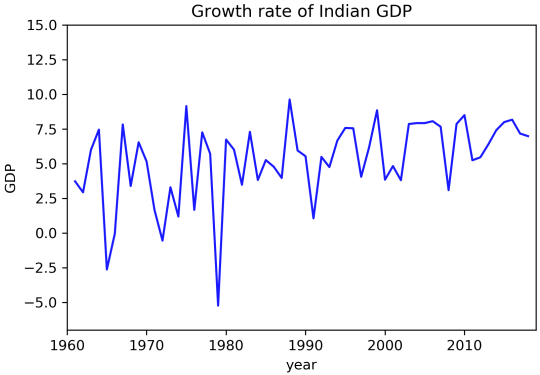


Figure: Growth rate of Indian GDP.

Quantity of currency (held by public) time series

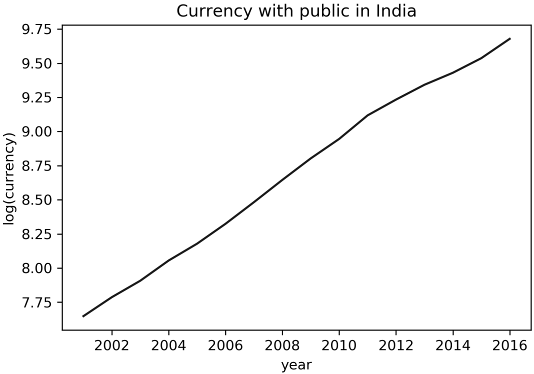


Figure: Quantity of currency held by public in India.

Reliance share price time series

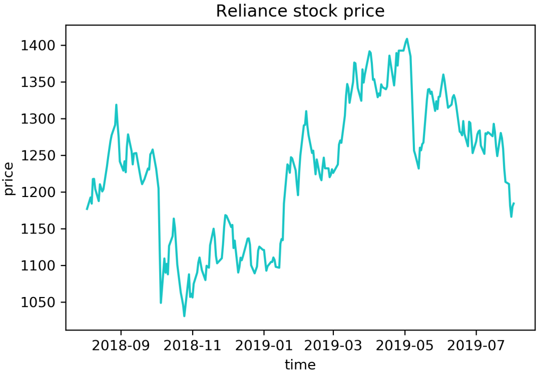


Figure: Reliance share price evoution.

Reliance volume traded time series

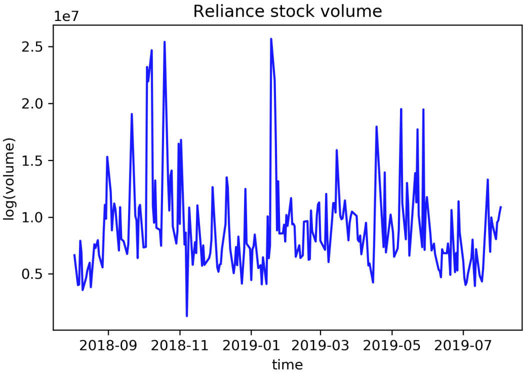


Figure: Reliance volume traded evoution.

Correlation $\neq$ Causation!

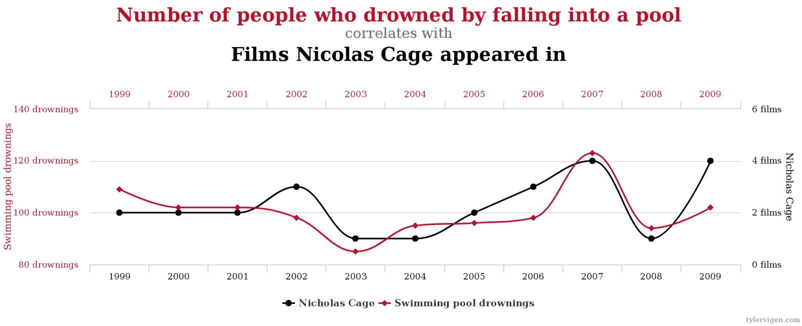


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Correlation $\neq$ Causation!

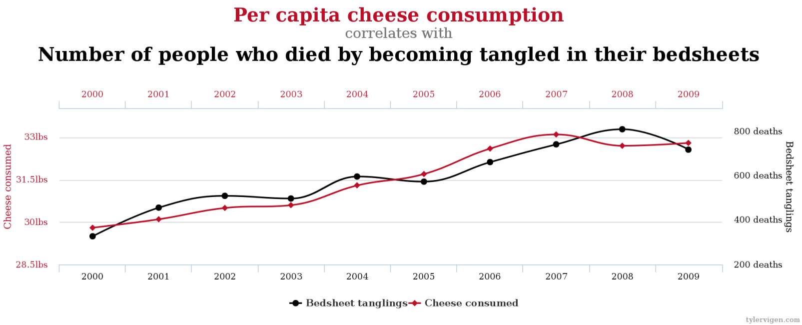


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Correlation $\neq$ Causation!

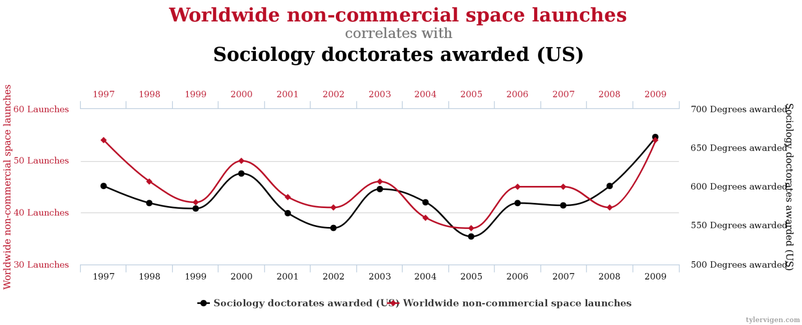


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Correlation $\neq$ Causation!

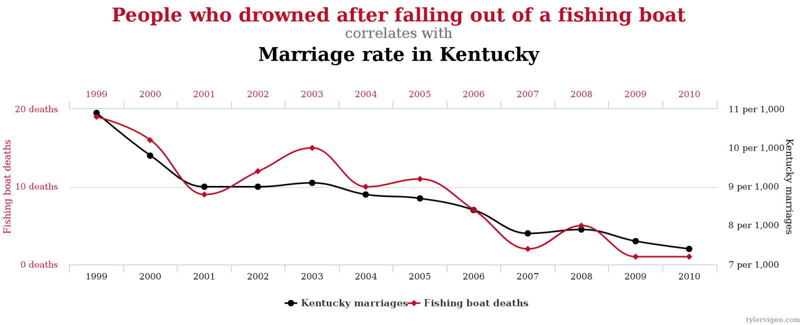


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Correlation $\neq$ Causation!

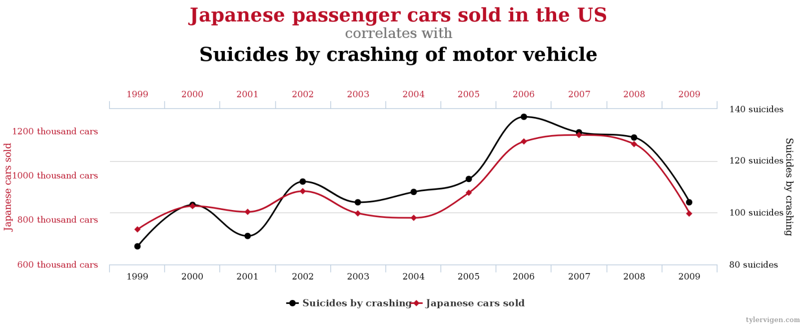


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Correlation $\neq$ Causation!

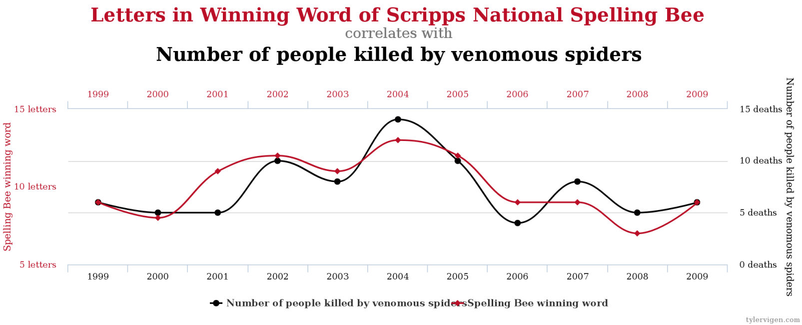


Figure: Figure taken from https://www.tylervigen.com/spurious-correlations compiled by Tyler Vigen.

Properties

Main observed properties of time series:

- Trend
- Fluctuations
- Seasonal variations

We model the time-series (with or without trend) with random variables.

Some definitions

We have N observations: $y_1,\ y_2\ ...y_N$.

Say, we have constructed an estimate of some population-specific parameters (vector). let us call it $\hat{\theta}$.

Goal

We want to know how correct is $\hat{\theta}$ as a description of the true data generating process with parameters θ.

The literature has developed asymptotic theory i.e, $N \rightarrow \infty$.

Persistence and volatility

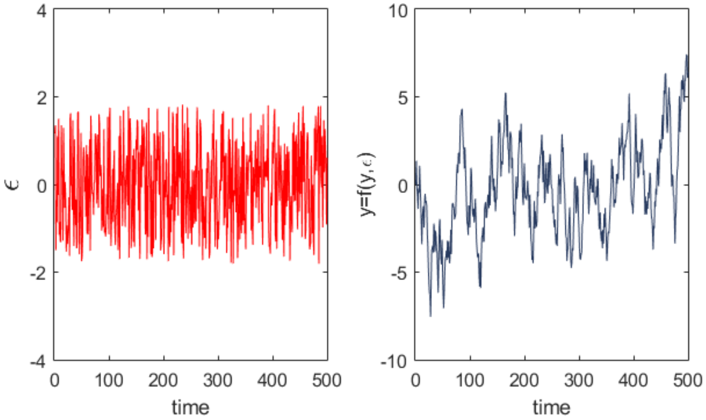


Figure: What is the difference between these two figures?

Strong stationarity

Strong stationarity

A stochastic process $\{x_t\}$ is *strongly* stationary if the joint probability distribution function of $\{x_{t-\tau}, \ldots, x_t, \ldots, x_{t+\tau}\}$ is independent of the time point t for all τ.

Weak stationarity

Weak stationarity

A stochastic process $\{x_t\}$ is *weakly* stationary if the first two moments $E(x_t)$ and $E(x_t^2)$ are finite and the lagged correlation $E(x_t x_\tau)$ is finite and depends only on the lag τ.

Such processes are also called *covariance stationary* processes.

White Noise

The building block of time series models is the white noise process. Let us assume:

$$\varepsilon_t \sim \text{i.i.d } N(0, \sigma_\varepsilon^2). \tag{1}$$

Then, the implications of this assumption would be:

- $E(\varepsilon_t) = E(\varepsilon_t | \varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}) = 0$
- $E(\varepsilon_t \varepsilon_{t-j}) = Cov(\varepsilon_t \varepsilon_{t-j}) = 0$
- $Var(\varepsilon_t) = Var(\varepsilon_t | \varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}) = \sigma_\varepsilon^2$

White Noise

Basic ideas:

- White noise
- Lack of serial correlation
- Conditional homoskedasticity

Examples

Example of data: Quantities which non-trivially depend on their own history.

- GDP growth rates.
- firm size growth rates.
- temperature.
- ...

How to build an ARMA model?

Class of models created by taking linear combinations of white noise.

- $AR(1)$:
$$x_t = \phi x_{t-1} + \varepsilon_t \tag{2}$$
- $MA(1)$:
$$x_t = \theta \varepsilon_{t-1} + \varepsilon_t \tag{3}$$
- $AR(p)$:
$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \tag{4}$$
- $MA(q)$:
$$x_t = \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \tag{5}$$

How to build an ARMA model?

Most general form:

- $ARMA(p, q)$:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{6}$$

Without loss of generalization, we assume that $\bar{x} = 0$. If required, we can always introduce a constant term in the following way:

$$x_t - \bar{x} = \phi(x_{t-1} - \bar{x}) + \varepsilon_t \tag{7}$$

which again follows $AR(1)$ process with a constant. Nothing changes fundamentally.

Choice of representation

We have seen two representations, AR and MA. Which one is more desirable?

- There is mathematical rule. It's more about convenience.
- For finding unconditional moments, MA process is desirable.
- If we need a representation of dependence on past values (which is more intuitive; e.g. higher gdp growth leads to higher gdp growth), then AR process is desirable.

$AR(1)$ to $MA(\infty)$ by recursive substitution

Here, we show that one can go from one representation to the other very easily. Let's say

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \text{where } |\phi| < 1. \tag{8}$$

Using lag operator, we can expand on the expression:

$$\begin{aligned} (1-\phi L)x_t &= \varepsilon_t \\ x_t &= \frac{\varepsilon_t}{(1-\phi L)} \\ &= (1+\phi L+\phi^2 L^2+\phi^3 L^3+\ldots)\varepsilon_t \\ &= \varepsilon_t+\phi\varepsilon_{t-1}+\phi^2\varepsilon_{t-2}+\ldots \\ &= \sum_{j=0}^{\infty}\phi^j\epsilon_{t-j}. \end{aligned} \tag{9}$$

$AR(1)$ to $MA(\infty)$ by recursive substitution

If the process started finite t periods back, we can expand it as:

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_t \\ &= \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \varepsilon_t \\ &= \phi^t x_0 + \sum_{j=0}^{t-1} \phi^j \varepsilon_{t-j}. \end{aligned} \tag{10}$$

If we assume that the process started infinite periods ago, then we have

$$x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \tag{11}$$

Example: Expanding series with lag polynomials

Let's consider an $AR(2)$ process given by

$$x_t = \phi_2 x_{t-2} + \phi_1 x_{t-1} + \varepsilon_t. \tag{12}$$

This can be rewritten as

$$(1 - \phi_1 L - \phi_2 L^2) x_t = \varepsilon_t \tag{13}$$

which in turn can be expressed as (assuming invertibility)

$$x_t = \frac{\varepsilon_t}{(1 - \phi_1 L - \phi_2 L^2)}. \tag{14}$$

Representation

ACF is a fundamental representation of the ARMA process.

- Note that ARMA description is not unique as we can switch between AR, MA and ARMA representation.
- Since the autocorrelation function is fundamental, we look for a representation of that.
- Not any set of numbers $1, \rho_1, \rho_2, \ldots$, will represent a valid ACF.

Autocovariance

We will denote autocovariance by

$$\gamma_j = Cov(x_t, x_{t-j}). \tag{15}$$

Note that here the time index t doesn't matter as covariance across j time points will be the same for all time points t.
The j-lag autocovariance can be written as

$$\gamma_j = E(x_t x_{t-j}). \tag{16}$$

Autocovariance

Note that 0-lag autocorrelation is just the variance:

$$\gamma_0 = Var(x_t). \tag{17}$$

Autocorrelation function

Now we can define the autocorrelation function (a.c.f.) as

$$\rho_j = \frac{\gamma_j}{\gamma_0}. \tag{18}$$

Example: ACF for AR(1) process

AR(1):

Consider a process $x_t = \phi x_{t-1} + \varepsilon_t$. Therefore,

$$\gamma_0 = \mathrm{var}(x_t) = \frac{\sigma_\epsilon^2}{1-\phi} \tag{19}$$

$$\gamma_1 = E(x_t x_{t-1}) = \frac{(\phi\sigma_\epsilon^2)}{(1-\phi)} = \phi\gamma_0 \tag{20}$$

$$\gamma_2 = E((\phi x_{t-1} + \varepsilon_t)x_{t-2}) = \phi^2 E(x_{t-2}^2) = \phi^2\gamma_0 \tag{21}$$

Example: ACF for AR(1) process

We can easily derive the autocorrelation function as

$$\begin{aligned} \rho_1 &= \phi && (22) \\ \rho_2 &= \phi^2 && (23) \\ &\vdots \\ \rho_j &= \phi^j \ \ \forall\, j > 0. && (24) \end{aligned}$$

Wold Decomposition Theorem

Theorem

Any mean zero covariance stationary process x_t can be represented in the form:

$$x_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} + \eta_t \tag{25}$$

1. $\epsilon_t \equiv x_t - P(x_t | x_{t-1}, x_{t-2}, ...)$
2. $P(\epsilon_t | x_{t-1}, x_{t-2}, ...) = 0$, $E(\epsilon_t x_{t-j}) = 0$, $E(\epsilon_t = 0)$, $E(\epsilon_t^2) = \sigma_\epsilon^2$, $E(\epsilon_t \epsilon_s) = 0$ *for all* $t \neq s$
3. *All the roots of $\theta(L)$ are on or outside the unit circle, i.e. (unless there is a unit root) the MA polynomial is invertible.*

Theorem

Any mean zero covariance stationary process x_t can be represented in the form:

$$x_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} + \eta_t \tag{26}$$

4. $\sum_{j=0}^{\infty} \theta_j^2 < \infty$, $\theta_0 = 1$
5. $\{\theta_j\}$ *and* $\{\epsilon_s\}$ *are unique.*
6. η_t *is linearly deterministic, i.e.* $\eta_t = P(\eta_t | x_{t-1}, ..)$

Properties of ACF and PACF

Process	ACF ($\rho(h)$)	PACF ($\pi(h)$)
AR(p)	infinite; dampened exponential or sinusoidal waves	finite; $\pi(h) = 0$ for $h > p$
MA(q)	finite; $\rho(h) = 0$ for $h > q$	infinite dampened exponential or sinusoidal waves
ARMA(p, q)	as AR(p) for $h > q$	as MA(q) for $h > p$

Estimation of ARMA process

Goal:

We have a series of observations $X_1, \ X_2, \ \ldots, X_T$. We want to fit an ARMA($p, \ q$) to these observations.

we rely on a fundamental result.

Basic framework

A stationary data generating process can be approximated by ARMA($p, \ q$) process.

Estimation of ARMA process

What we would like to do:

- First, we need to determine the order (p and q) of the process.
- Then we need to find the values of the coefficients.
- Also, need to identify the error term variance.

Box-Jenkins methodology (1976)

The basic recursive model selection technique works as follows:

- Identify the model
- Estimate the coefficients or parameters.
- Perform diagnostics of the fitted model.
- If the model fails the tests, go to the first step.
- If the model seems satisfactory, one can make predictions.

Model identification

- First, check stationarity of data $X_1, X_2, \ldots, X_T$.
- Visual inspection.
- Statistical tests for stationarity (Dickey-Fuller, Phillips-Perron etc.).
- Transform the data to achieve stationarity.
  - For example, GDP per capita (G_t) can be non-stationary but the growth rate ($g_t = log(G_t) - log(G_{t-1})$) can be stationary.
  - Sometimes, more than one round of difference operator can be required to achieve stationarity.
  - There is an idea of *fractional difference* as well (we will skip the discussion here).

Model identification

The net step is to select the orders p and q.

- Compute empirical ACF and PACF (software will do it for you).
- This gives you an idea about MA and AR order, respectively.
- One can utilize more sophisticated statistical criteria as well.

Estimate the parameter vector

Once we fix the orders, we know the number of parameters to be estimated.

- Ordinary least square can be applied.
- Maximum likelihood estimation can be applied.

Perform diagnostics on the model and use it for prediction

next, we check for autocorrelation in the residuals.

- Ljung-Box test.
- If the residuals do not have any autocorrelation, then the model is good.
- Else, we need to modify.

If the model is good to go,

- we can use it for prediction.

Estimation of the lag orders

There are obviously two errors that we have to be careful about.

- We can pick large p and q: Model overfits the data $\rightarrow$ High in-sample fit but poor out-of sample performance.
- We can pick small p and q: Model underfits the data $\rightarrow$ Low in-sample fit but better out-of sample performance.

Either way it is bad since the maximum likelihood estimator would not be consistent.

- We can use ACF and PACF. But such visual inspection may not be very accurate.
- We can use information criteria.

Information criteria

This is a standard tool for model selection. We want to minimize the information criterion to come up with a better model.

- If we over-parameterize, then the in-sample fit has to necessarily increase.
- Let us denote the fit of the model by the variance of the residuals ($\hat{\sigma}^2_{p,q}$).
- We penalize higher values of p and q and correct for $\hat{\sigma}^2_{p,q}$.

Information criteria (standard procedures)

Main objective: Minimize information criterion IC with respect to p and q where

- Akaike IC $AIC(p,q) = \log(\hat{\sigma}^2_{p,q}) + \frac{2}{T}\Big(p+q\Big)$.
- Bayesian or Schwarz IC $BIC(p,q) = \log(\hat{\sigma}^2_{p,q}) + \frac{\log(T)}{T}\Big(p+q\Big)$.
- Hannan-Quinn IC $HQIC(p,q)$: $\log(\hat{\sigma}^2_{p,q}) + \frac{2(\log(\log(T)))}{T}\Big(p+q\Big)$.

Information criteria (standard procedures)

General rule:

BIC and HQIC penalizes over-parameterization more than AIC and hence typically economizes on the number of parameters.

However, AIC might pick a better model (closer to the true data generating process) in small samples.

Rule of thumb:

You can use all of them. None of these clearly dominates others. AIC and BIC are more popular than HQIC.

Estimation of AR(p) process

Let us say we have a model of the form:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \tag{27}$$

- We want to estimate $\phi_1,\ \phi_2, ...\phi_p$.
- We can use Yule-Walker estimators.

However, here we discuss more standard OLS procedure.

Estimation of AR(p) process

Consider an equation

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \tag{28}$$

- Say, x_t is the endogenous variable,
- $x_{t-1},\ x_{t-2}, ...x_{t-p}$ are exogenous variables,
- $\phi_1,\ \phi_2, ...\phi_p$ are coefficients,
- ϵ_t is the error term.

Estimation of AR(p) process

In matrix form, we can write

$$\begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_T \end{bmatrix} = \begin{bmatrix} 1 & X_p & X_{p-1} & X_{p-2} & \dots & X_1 \\ 1 & X_{p+1} & X_p & X_{p-1} & \dots & X_2 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{T-1} & X_{T-2} & X_{T-3} & \dots & X_{T-p} \end{bmatrix} \times \begin{bmatrix} c \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} u_{p+1} \\ u_{p+2} \\ \vdots \\ u_T \end{bmatrix} \tag{29}$$

Let us write it in a more convenient form

$$Y = X\beta + \epsilon. \tag{30}$$

Estimation of AR(p) process

Clearly the OLS estimator is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y. \tag{31}$$

Now we may estimate the error variance by finding the residuals

$$\hat{\epsilon} = Y - X\hat{\beta}_{OLS}, \tag{32}$$

and construct

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{T-p}. \tag{33}$$

Estimation of AR(p) process

Some technical issues:

- regressors are correlated with error terms
- the origin of the observation $[X_1,\ X_2, \ldots,\ X_p]$ affect the OLS estimates.

Result

For AR(p) models, $\hat{\beta}_{OLS}$ is consistent and asymtotically efficient.

Note

For ARMA($p,\ q$) models, $\hat{\beta}_{OLS}$ can not be estimated since the error terms are not observable.

Estimation of ARMA(p, q) process

We utilize maximum likelihood estimation.

- Consistent.
- Asymptotically efficient.
- Asymptotically Normally distributed.
- Robust against deviation from assumption of normality of the data-generating process.

Estimation of ARMA(p, q) process

Let us say we have a model of the form:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{34}$$

with normally distributed white noise ϵ_t i.e., $\epsilon_t \sim N(0, \sigma^2)$.

Log-likelihood function: Consider the starting values

$$\begin{aligned} X_0 &\equiv [x_0 \; x_{-1} \; \ldots \; x_{-p+1}]' \\ \epsilon_0 &\equiv [\epsilon_0 \; \epsilon_{-1} \; \ldots \; \epsilon_{-p+1}]'. \end{aligned} \tag{35}$$

Estimation of ARMA(p, q) process

Exact Gaussian likelihood of $x = (x_1,\ x_2, \ldots,\ x_T)'$ is given by

$$L(\beta|x) = (2\pi)^{-T/2}|\Gamma(\beta)|^{-1/2} exp\left(-\frac{1}{2}x'\Gamma(\beta)^{-1}x\right) \tag{36}$$

where $\Gamma(\beta) = E(xx')$ is the $T \times T$ covariance matrix of x which is a function of β.

Estimation of ARMA(p, q) process

The conditional loglikelihood function is given by

$$\mathcal{L}(\beta|x_0, \epsilon_0) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T}\frac{\epsilon_t^2}{2\sigma^2}. \qquad (37)$$

- $\mathcal{L}(\beta|x_0, \epsilon_0)$ is a nonlinear function of β.
- We resort to numerical techniques to maximize the likelihood function.
- All softwares/programming languages (R, matlab, python, stata etc) have these kind of programs inbuilt.

Vector autoregression model

Vector autoregression models are objects that follows the following structure:

$$\begin{aligned} X_{1t} &= \phi^1_{11}X_{1,t-1} + \phi^1_{12}X_{1,t-2} + \ldots + \phi^1_{21}X_{2,t-1} + \ldots + \epsilon_{1t} \\ X_{2t} &= \phi^2_{11}X_{1,t-1} + \phi^2_{12}X_{1,t-2} + \ldots + \phi^2_{21}X_{2,t-1} + \ldots + \epsilon_{2t} \\ &\vdots \\ X_{nt} &= \phi^n_{11}X_{1,t-1} + \phi^n_{12}X_{1,t-2} + \ldots + \phi^n_{21}X_{2,t-1} + \ldots + \epsilon_{nt}. \end{aligned} \tag{38}$$

Why VAR representation is useful?

The idea is that any ARMA(p,q) process can be projected in to an AR(1) process which is essentially VAR. For an example, consider the following ARMA(2,1) process:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{39}$$

$$\begin{pmatrix} X_t \\ X_{t-1} \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \epsilon_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} [\epsilon_t]. \tag{40}$$

Why VAR representation is useful?

This expression can be written as:

$$x_t = \Theta x_{t-1} + \Gamma w_t \tag{41}$$

where

$$\Gamma = \begin{pmatrix} \sigma_\epsilon \\ 0 \\ \sigma_\epsilon \end{pmatrix} \tag{42}$$

and w_t captures the normalized noise with $E(w_t w_t') = I$.

Granger Causality

Causality: If an event A takes place regularly after another event B, then the preceding event may cause the event that follows.

Definition

x_{1t} Granger causes x_{2t} if x_{1t} has a predictive component for x_{2t}, given past relaizations of x_{2t}.

Granger Causality

Consider a bivariate VAR:

$$\begin{aligned} x_{1t} &= \theta_{11}(L)x_{1,t-1} + \theta_{12}(L)x_{2,t-1} + \epsilon_{1t} \\ x_{2t} &= \theta_{21}(L)x_{1,t-1} + \theta_{22}(L)x_{2,t-1} + \epsilon_{2t}. \end{aligned} \tag{43}$$

Granger Causality

Suppose x_1 Granger causes x_2, but not the other way round. Then

$$\begin{aligned} x_{1t} &= \theta_{11}(L)x_{1,t-1} + \qquad\qquad\qquad\quad \epsilon_{1t} \\ x_{2t} &= \theta_{21}(L)x_{1,t-1} + \theta_{22}(L)x_{2,t-1} + \epsilon_{2t}. \end{aligned} \tag{44}$$

Summary of predictive algorithm

Steps:

- Given an $ARMA(p, q)$ system, convert it into a $VAR(1)$ model:

$$x_t = \Theta x_{t-1} + \Gamma w_t. \tag{45}$$

- Use the following formulae:

$$E_t(x_{t+\tau}) = \Theta^\tau x_t \tag{46}$$

and

$$var_t(x_{t+\tau}) = \sum_{j=0}^{\tau-1} \Theta^j \Gamma\Gamma'(\Theta^j)'. \tag{47}$$

Impulse Response function

IRF is a very intuitive and easy way to understand the structure of a VAR model.

- It captures how shock propagates from one variable to another.
- Provides an idea about how shocks diffues over time.
- Useful for model validation (e.g. DSGE models).

Impulse response function: Basic idea

Consider a simple AR(1) process:

$$x_t = \phi x_{t-1} + \epsilon_t. \tag{48}$$

Imagine that $x_0 = 0$ and ϵ_t is given an unit shock. Then x responds as the following:

$$\begin{array}{rccccl} t: & 1 & 2 & 3 & 4\ldots \\ \epsilon_t: & 1 & 0 & 0 & 0\ldots \\ x_t: & 1 & \phi & \phi^2 & \phi^3\ldots \end{array} \tag{49}$$

Impulse response function: Basic idea

Note that by inversion, we get

$$x_t = (1 + \phi L + ... + \phi^n L^n + ...)\epsilon_t. \tag{50}$$

Therefore, the series of MA coefficients constitute the impulse response function.

IRF

The coefficients of MA representation of an ARMA process constitutes the corresponding IRF.

Impulse response function: Basic idea

As an example, consider an MA(2) process:

$$x_t = (1 + \gamma_1 L + \gamma_2 L^2)\epsilon_t. \tag{51}$$

The IRF is:

$$\begin{array}{rcccc} t: & 1 & 2 & 3 & 4\dots \\ \epsilon_t: & 1 & 0 & 0 & 0\dots \\ x_t: & 1 & \gamma_1 & \gamma_2 & 0\dots \end{array} \tag{52}$$

Impulse response function

Generally, consider a VAR(1) process:

$$X_t = \Theta X_{t-1} + \Gamma \epsilon_t. \tag{53}$$

Then the impulse response function is given as

$$\begin{array}{rcccc} t: & 1 & 2 & 3 & 4\ldots \\ \epsilon_t: & 1 & 0 & 0 & 0\ldots \\ X_t: & \Gamma & \Theta\Gamma & \Theta^2\Gamma & \Theta^3\Gamma\ldots \end{array} \tag{54}$$

Problems of impulse response function

They are generally non-unique. Not part of the present discussion.

More on VAR

VAR models are quite useful.

- They are agnostic to the theory.
- Provides a summary of lagged correlation between multiple variables at one go.
- Can provide an idea about causality (in the *Granger* sense).
- We can use theoretical restrictions to identify VAR models.

Modeling the second moment

We will introduce the GARCH class of models.

What does financial data look like?

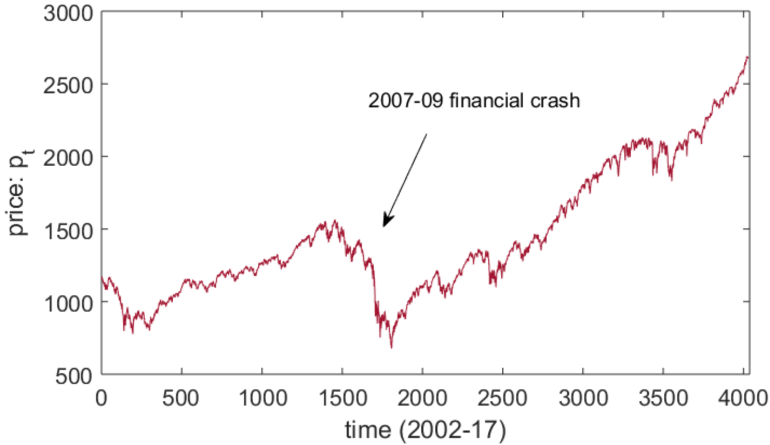


Figure: S&P 500 index. Growing trend with occasional downswings.

Do we care about price or return?

Return is the most important factor.

$$r_t = log(p_t) - log(p_{t-1}). \tag{55}$$

What does financial data look like?

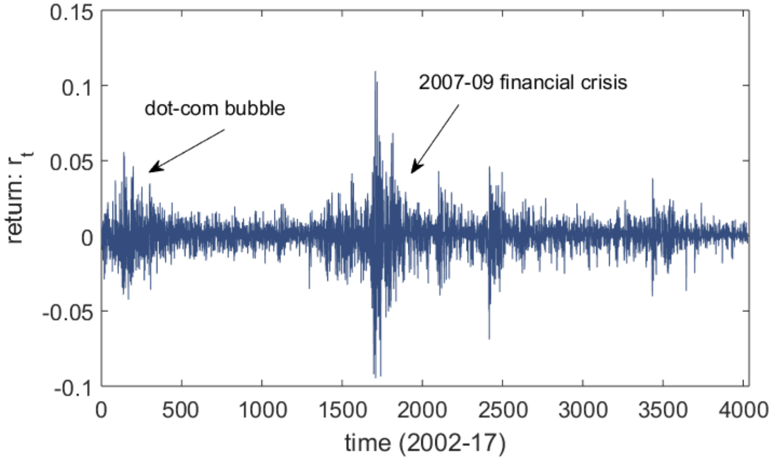


Figure: S&P 500 index fluctuations.

Some interesting properties

What can we read from the data?

- Lots of movement! Wild swings are observed.
- Average return is close to zero.
- level of return seems to have no relationship over time. A good return today does not indicate a good return tomorrow.
- Volatile periods tend to cluster.
- Market has ‘memory’ in volatility!

How to measure 'memory'?

We need some tools.

Definition (Autocorrelation)

Autocorrelation (or serial correlation) is the correlation of a time series with a delayed copy of itself as a function of delay (also called *lag*).

Formally, the expression is

$$R(\tau) = \frac{E\left((X_t - \mu)(X_{t+\tau} - \mu)\right)}{\sigma^2}. \tag{56}$$

Intuition: It is just like cross-correlation!

What does financial data look like?

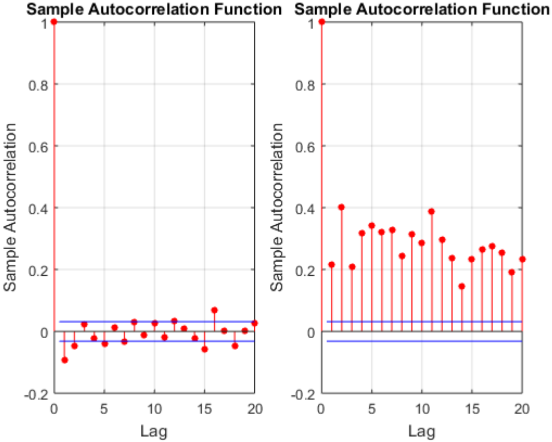


Figure: S&P 500 index: Autocorrelations (left: r_t, right: r_t^2).

How to make sense of it?

We have a model called GARCH (Generalized Autoregressive Conditional Heteroscedastic) that allows you to find out how volatile a market is.

A simple example of GARCH(1,1) is as follows:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha \sigma_{t-1}^2 + \beta r_{t-1}^2. \end{aligned} \tag{57}$$

where ϵ_t is an independent standard normal random variable.

What does financial data look like?

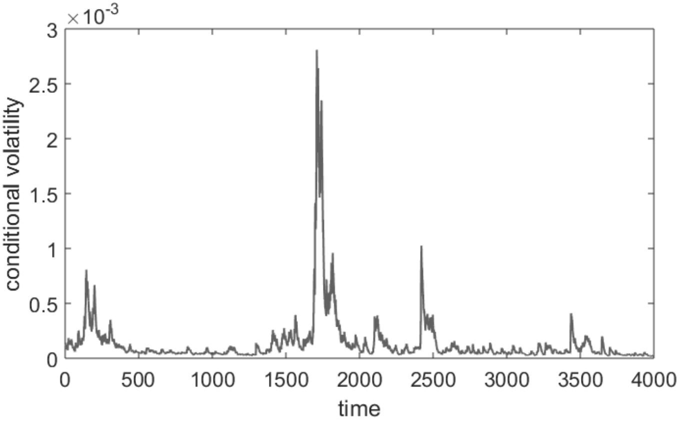


Figure: S&P 500 index: Latent volatility.

GARCH(p,q) process

Let us define the following process:

$$\begin{aligned} y_k &= \sigma_k \varepsilon_k \\ \sigma_k^2 &= w + \sum_{i=1}^{p} \alpha_i y_{k-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{k-j}^2 \end{aligned} \tag{58}$$

We assume that

- $w > 0$
- $\alpha_i \geq 0$
- $\beta_i \geq 0$

Note:

No structural reason behind the above model. But you will see that it is very useful.

GARCH(p, q) process

Other assumptions:

- $\{\varepsilon_i\}$ is IID, $E(\varepsilon_0) = 0$, $E(\varepsilon_0^2) = 1$.
- $(\sum_i^p \alpha_i + \sum_j^q \beta_j) < 1$ for uniqueness and stationarity.

Question

Why stationarity even if fluctuations depend on the level of volatility time dependent?

Stationarity

Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates. *(Thomson 1994)*

Thomson, D.J. 1994. Jackknifing multiple-window spectra. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, VI, 73-76

Recall the Definitions

Definition (Strong stationarity:)

A strictly/strongly stationary stochastic process is one where the joint statistical distribution of $X_{t_1}, \ldots, X_{t_l}$ is the same as the joint statistical distribution of $X_{t_1+\tau}, \ldots, X_{t_l+\tau}$ for all l and τ.

Definition (Weak stationarity:)

A weakly stationary stochastic process is one of which

- the mean $E(x_t) = c$ where c is independent of time,
- the variance $var(x_t) = \sigma^2$ where σ is independent of time,
- and finally, the covariance $cov(x_t, x_{t-l}) = \gamma(l)$ i.e., it depends only on the lag l.

Wold Decomposition Theorem

Theorem

Any mean zero covariance stationary process x_t can be represented in the form:

$$x_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} + \eta_t \tag{59}$$

1. $\epsilon_t \equiv x_t - P(x_t | x_{t-1}, x_{t-2}, ...)$
2. $P(\epsilon_t | x_{t-1}, x_{t-2}, ...) = 0$, $E(\epsilon_t x_{t-j}) = 0$, $E(\epsilon_t = 0)$, $E(\epsilon_t^2) = \sigma_\epsilon^2$, $E(\epsilon_t \epsilon_s) = 0$ *for all* $t \neq s$
3. *All the roots of $\theta(L)$ are on or outside the unit circle, i.e. (unless there is a unit root) the MA polynomial is invertible.*

Theorem

Any mean zero covariance stationary process xt can be represented in the form:

$$x_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j} + \eta_t \tag{60}$$

4. $\sum_{j=0}^{\infty} \theta_j^2 < \infty$, $\theta_0 = 1$
5. $\{\theta_j\}$ *and* $\{\epsilon_s\}$ *are unique.*
6. η_t *is linearly deterministic, i.e.* $\eta_t = P(\eta_t | x_{t-1}, ..)$

Non-stationarity

Non-stationarity might arise in the following ways.

- Deterministic trend (or trend stationarity)
- Regime shift in level
- Change in variance
- Unit roots (stochastic trend)

Trend stationarity

Definition (Trend stationarity:)

A process $\{X_t\}$ is called trend stationary if

$$X_t = f(t) + \epsilon_t \tag{61}$$

where ϵ_t is a stationary process.

Trend stationarity

Example:

Let us say z_t is stationary AR(1) process

$$z_t = \rho z_{t-1} + \epsilon_t. \tag{62}$$

Let us assume that process x_t has the following form

$$x_t = a + b.t + c.t^2 + z_t. \tag{63}$$

Then x_t is a trend stationary process.

Difference stationarity

Definition (Difference stationarity:)

A process x_t is called difference stationary if

$$\Delta^p x_t = \epsilon_t \tag{64}$$

where ϵ_t is stationary for some p.

AR process with unit root

Consider the following process:

$$x_t = x_{t-1} + \epsilon_t. \tag{65}$$

The characteristic polynomial is $\theta(z) = 1 - z$. Therefore, it has a unit root, $\theta(1) = 0$.

By backward substitution, we get

$$x_t = x_0 + \epsilon_1 + \ldots + \epsilon_t. \tag{66}$$

Important point:

Effect of any shock remains. Property of mean reversion is absent.

AR process with unit root

From the above expression,

$$var(x_t) = t\sigma^2. \tag{67}$$

Also, the covariance is given by (assume $t \geq s$)

$$\begin{aligned} E((x_t - x_0)(x_{t-s} - x_0)|x_0) &= E((\epsilon_1 + \epsilon_2 + \ldots + \epsilon_t)(\epsilon_1 + \epsilon_2 + \ldots + \epsilon_{t-s})) \\ &= (t-s)\sigma^2. \end{aligned}$$

Autocorrelation is given by

$$\begin{aligned} \rho(x_t, x_{t-s}|x_0) &= \frac{cov(x_t, x_{t-s}|x_0)}{\sqrt{V(x_t|x_0)V(x_{t-s}|x_0)}} \\ &= \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2(t-s)\sigma^2}} \\ &= \sqrt{\frac{t-s}{t}}. \end{aligned} \tag{69}$$

Random walk with a drift

Consider a process:

$$x_t = \delta + x_{t-1} + \epsilon_t. \tag{70}$$

By backward substitution, we get

$$x_t = x_0 + \delta t + \sum_{j=1}^{t} \epsilon_t. \tag{71}$$

Unit root testing

Estimate an autoregressive model and test whether there is an unit root or not.

Note:

The asymptotic distribution for a unit root test is non-standard. It does not converge to normal distribution.

Dickey-Fuller test

Consider an AR(1) model

$$x_t = \rho x_{t-1} + \epsilon_t. \tag{72}$$

Hypothesis test is conducted on

$$H_0 : \quad \rho = 1 \quad \text{against} \quad H_1 : \quad \rho \in (-1, 1). \tag{73}$$

Dicky-Fuller test:

The DF test statistic is simply the $t - ratio$

$$\hat{t} = \frac{\hat{\rho} - 1}{se(\hat{\rho})}. \tag{74}$$

Dicky-Fuller distribution is simulated (and it is not normally distributed).

Dickey-Fuller test

Note that the process

$$x_t = \rho x_{t-1} + \epsilon_t. \tag{75}$$

can be written as

$$\Delta x_t = \pi x_{t-1} + \epsilon_t. \tag{76}$$

where $\pi = \rho - 1$.

Therefore, the hypothesis test can also be conducted on

$$H_0: \quad \pi = 0 \quad \text{against} \quad H_1: \quad \pi \in (-2, 0). \tag{77}$$

The DF test statistic in this case is

$$\hat{t} = \frac{\hat{\pi}}{se(\hat{\pi})}. \tag{78}$$

Augmented Dickey-Fuller test

Suppose we have an AR(p) process:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \ldots + \rho_p x_{t-p} + \epsilon_t. \tag{79}$$

This can be rewritten as

$$\Delta x_t = \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t \tag{80}$$

where $\pi,\ \theta_1,\ \theta_2,\ \ldots \theta_{p-1}$ are defined suitably. Therefore, the hypothesis test can also be conducted on

$$H_0: \quad \pi = 0 \quad \text{against} \quad H_1: \quad \pi \in (-2, 0). \tag{81}$$

The DF test statistic in this case is

$$\hat{t} = \frac{\hat{\pi}}{se(\hat{\pi})}. \tag{82}$$

Augmented Dickey-Fuller test with a constant term

Suppose we have an AR(p) process:

$$\Delta x_t = \delta + \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t \tag{83}$$

We can still check

$$H_0 : \quad \pi = 0 \quad \text{against} \quad H_1 : \quad \pi \in (-2, 0) \tag{84}$$

with same distribution of t-ratio. However, that incorporate a deterministic trend due to δ. A better option is to test

$$H_0 : \quad \pi = \delta = 0. \tag{85}$$

The joint hypothesis can be tested by a LR test,

$$LR(\pi = \delta = 0) \; = \; -2(\log \; L_0 - \log \; L_1) \tag{86}$$

where LR statistic follows DF_c^2 under the null hypothesis.

Augmented Dickey-Fuller test with a trend term

Suppose we have an AR(p) process:

$$\Delta x_t = \delta + \gamma t + \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t \tag{87}$$

We can still check

$$H_0: \quad \pi = 0 \quad \text{against} \quad H_1: \quad \pi \in (-2, 0) \tag{88}$$

with same distribution of t-ratio. A better option is to test

$$H_0: \quad \pi = \gamma = 0. \tag{89}$$

The joint hypothesis can be tested by a LR test,

$$LR(\pi = \delta = 0) \;=\; -2(\log \;\, L_0 - \log \;\, L_1) \tag{90}$$

where LR statistic follows DF_l^2 under the null hypothesis.

Summary of the tests (borrowed from Morten Nyboe Tabor's notes)

Model	DF	LR
$\Delta x_t = \pi x_{t-1} + \epsilon_t$	DF_l	DF_l^2
$\Delta x_t = \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t$	DF	DF^2
$\Delta x_t = \delta + \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t$	DF_c	DF_c^2
$\Delta x_t = \delta + \gamma t + \pi x_{t-1} + \theta_1 \Delta x_{t-1} + \ldots + \theta_{p-1} \Delta x_{t-p+1} + \epsilon_t$	DF_l	DF_l^2

Hypothesis testing:

- For the ADF test:

$$H_0: \quad \pi = 0 \quad \text{against} \quad H_1: \quad \pi \in (-2, 0) \tag{91}$$

- For the likelihood ratio (LR) test:

$$H_0: \quad \pi = 0 \quad \text{or} \quad H_0: \quad \pi = \delta = 0 \quad \text{or} \quad H_0: \quad \pi = \gamma = 0. \tag{92}$$

Other tests

There is a host of other tests that you can use.

- Phillips-Perron test and its variants/modifications.

So far we have talked only about testing for unit roots.

KPSS test

This test enables us to check for stationarity.

Challenges and pitfalls

Unit Roots in Real GNP: Do We Know, and Do We Care?
Lawrence J. Christiano, Martin Eichenbaum
NBER Working Paper No. 3130, Issued in October 1989

Abstract: No, and maybe not.

Challenges and pitfalls

Power is the probability of rejecting the null hypothesis when it is incorrect.

	Null hypothesis	
Decision	True	False
Reject null	type I	power
Cannot reject null	correct	type II

Challenges and pitfalls

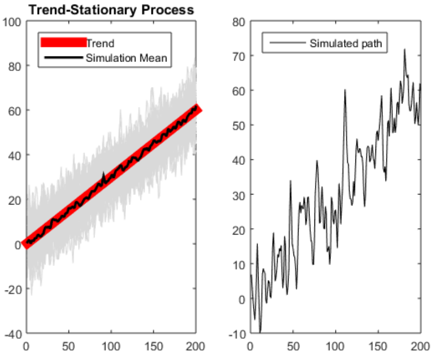


Figure: Simulation of trend-stationary processes

Challenges and pitfalls

It is extremely difficult to differentiate between trend stationary and difference stationary series over short time scale.

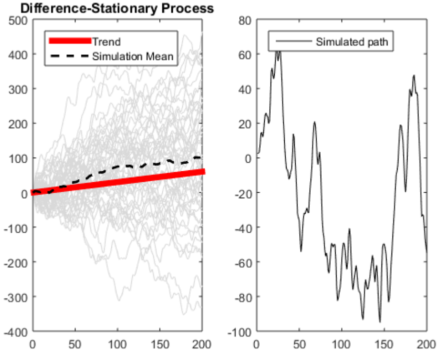


Figure: Simulation of difference-stationary processes

Challenges and pitfalls

It is extremely difficult to differentiate between trend stationary and difference stationary series over short time scale.

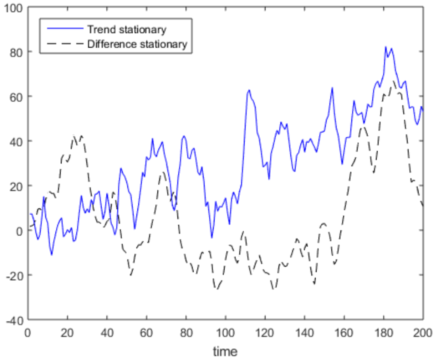


Figure: Comparison of simulated paths of trend-stationary and difference stationary processes.

Challenges and pitfalls

In the earlier figure we assumed a difference stationary process as follows:

$$\Delta y_t = 0.5 + e_t + 1.4e_{t-1} + 0.8e_{t-2} \tag{93}$$

and trend stationary process as follows:

$$y_t = 0.3t + 0.8y_{t-1} + e_t + 0.4e_{t-1}. \tag{94}$$

Replication

I have used rng(default) in matlab.

USD-INR exchange rate

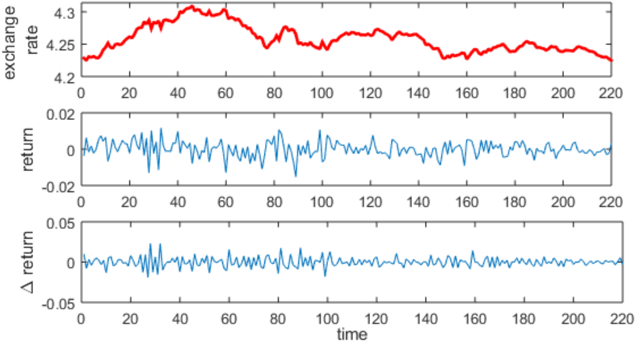


Figure: USD-INR exchange rate from 1st August, 2018 to 31st July, 2019. Panel(a): Log exchange rate; (b) log-return; (c) Δ log-return.

USD-INR exchange rate

Code:

$>> h = adftest(Exchange\_rate\_US)$

$>> h = 0$ i.e., this test fails to reject the null hypothesis of a unit root against the autoregressive alternative.

$>> h = adftest(Exchange\_rate\_US,' model',' ARD',' lags', 0 : 2)$

$>> h = 0\ \ 0\ \ 0$ i.e., this test fails to reject the null hypothesis of a unit root against the alternative with three lags with a drift term.

$>> h = adftest(Exchange\_rate\_US,' model',' TS',' lags', 0 : 2)$

$>> h = 0\ \ 0\ \ 0$ i.e., this test fails to reject reject the null hypothesis of a unit root against the trend-stationary alternative with three lags.

USD-INR exchange rate

Code:

$>> h = adftest(return\_US)$

$>> h = 1$ i.e., this test rejects the null hypothesis of a unit root against the autoregressive alternative.

$>> h = adftest(return\_US, 'model', 'ARD', 'lags', 0:2)$

$>> h = 1\ \ 1\ \ 1$ i.e., this test rejects the null hypothesis of a unit root against the alternative with three lags with a drift term.

$>> h = adftest(return\_US, 'model', 'TS', 'lags', 0:2)$

$>> h = 1\ \ 1\ \ 1$ i.e., this test rejects reject the null hypothesis of a unit root against the trend-stationary alternative with three lags.

USD-INR exchange rate: Phillips-Perron test

Code:

$>>$
$[h\_basic\_PP, pValue\_PP, stat\_PP, cValue\_PP, reg\_PP] = pptest(data\_US)$

$>> h = 0$ i.e., this test fails to reject the null hypothesis of a unit root against the autoregressive alternative.

$>> pValue\_PP = 0.6681$

$>> [h\_basic\_PP, pValue\_PP, stat\_PP, cValue\_PP, reg\_PP] = pptest(return\_US)$

$>> h = 1$ i.e., this test rejects the null hypothesis of a unit root against the autoregressive alternative.

$>> pValue\_PP = 1.0000e - 03$

USD-INR exchange rate: KPSS test for stationarity

Code:

$>> [h\_basic\_kpss] = kpsstest(data\_US)$

$>> h = 1$ i.e., this test rejects the trend-stationary null in favor of the unit root alternative.

$>> [h\_return\_kpss] = kpsstest(return\_US)$

$>> h = 0$ i.e., this test fails to reject the trend-stationary null.

References

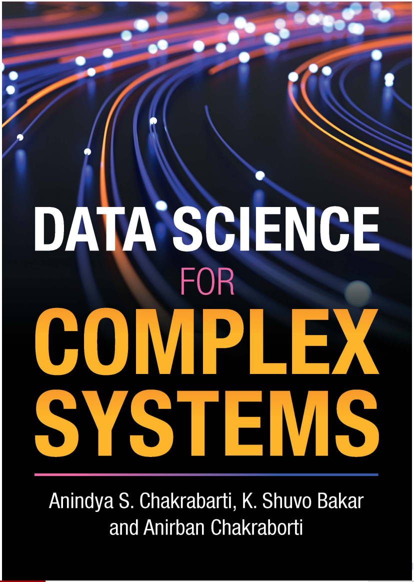
General references:

- J. Hamilton, *Time Series Analysis*, Princeton university press Princeton, NJ (1994)
- R. Tsay, *Analysis of Financial Time Series*, Wiley (2002)
- W. Enders, *Applied Econometric Time Series*, Wiley, Third edition (2013)
- P. Brockwell and R. Davis, *Introduction to Time Series and Forecasting*, Springer Texts in Statistics (2010)

Acknowledgement

This set of notes borrows extensively from the textbooks cited. These notes are meant only for teaching purpose.

Reference


DATA SCIENCE
FOR
COMPLEX
SYSTEMS
Anindya S. Chakrabarti, K. Shuvo Bakar
and Anirban Chakraborti