3 Linear Models (chapter 2 of the textbook)

In this course we ask: What? (what is the problem?) Now what? (what theories? methods? techniques?) So what? (your take away?)
Financial TS: Collection of a financial measurement over time (e.g., log return r_t), $\{r_1, r_2, ..., r_T\}$ (T data points).
Basic concepts: In order to use linear models to find what information is contained in $\{r_t\}$ (and use it to forecast future), we need the data to be "well behaved." Formally, the data need to be "stationary."

- Stationarity:
  - Strict stationarity: Distributions are time-invariant
  - Weak stationarity (or covariance stationarity): The first two moments are time-invariant

  Strict stationarity means that that the density function for any collection of times $\{r_{t_1}, r_{t_2}, ..., r_{t_m}\}$ is equal to the density function for a τ translation of the (for any arbitrary τ), $\{r_{t_1+\tau}, r_{t_2+\tau}, ..., r_{t_m+\tau}\}$. This is a very strong property, but it's a reasonable property to ask for if you are doing statistical inference (If you believe that the distribution is going to change over time, there is no benefit in studying and understanding it!). The strict stationarity requires knowledge about the entire density of the stochastic process. A weaker property is Weak stationarity or covariance stationarity where only the first two moments are time-invariant.

  What does weak stationarity mean in practice?

  Past: Time plot of $\{r_t\}$ varies around a fixed level within a finite range!

  Future: The first two moments of future r_t are the same as those of the data so that meaningful inferences can be made.

  Why do we care whether a series is stationary? First, stationary processes are better understood than non-stationary ones, and we know how to estimate them better. The test statistics of certain non-stationary processes do not follow the usual distributions. Further, if we regress two completely unrelated integrated processes on each other, then a problem called "spurious regression" can arise. In a nutshell, if X and Y are both trending, then regressing Y on X is likely to indicate a strong relationship between them, even though there is no real connection. They both depend upon time, so they would seem to be affecting each other.

  In general, the mean of a time-series $\{r_t\}$ is $E(r_t) = \mu_t$; the subscript denotes that the mean could depend upon the particular time. For example, if r_t is growing then its mean (or expected value) will also be growing. Tomorrow's r_{t+1} is expected to be greater than today's r_t. Likewise, the variance of r_t, denoted $Var(r_t)$ or σ_t^2, might depend upon the particular time period. For example, volatility might be increasing over time. More likely, volatility tomorrow might depend upon today's volatility.

  - $\{r_t\}$ is "mean stationary" if the expected value of $\{r_t\}$ at a particular time does not depend upon the particular time period in which it is observed. Thus, the unconditional expectation of $\{r_t\}$ is not a function of the time period t.

$$\mu = E[r_t]$$

- $\{r_t\}$ is said to be "variance stationary" if its variance is not a function of time.

$$Var[r_t] = \sigma^2 = E[(r_t - \mu)^2]$$

Figure 22 illustrates a time-series that is mean stationary (it reverts back to its average value) but is not variance stationary (its variance fluctuates over time with periods of high volatility and low volatility).

- $\{r_t\}$ is "covariance stationary" if the covariance of $\{r_t\}$ with its own lagged values depends only upon the length of the lag, but not on the specific time period nor on the direction of the lag.

Lag-k auto-covariance

$$\gamma_k = Cov(r_t, r_{t-k}) = E[(r_t - \mu)(r_{t-k} - \mu)].$$

Two important properties:

1. $\gamma_0 = \text{Var}[r_t]$
2. $\gamma_k = \gamma_{-k}$

Serial (or auto-) correlations

$$\rho_k = \frac{\text{Cov}(r_t, r_{t-k})}{\text{Var}[r_t]} = \frac{\gamma_k}{\gamma_0}$$

Auto-correlations refer to the way the observations in a time series are related to each other and is measured by the correlation between the current observation (r_t) and the observation ℓ period from the current period ($r_{t-\ell}$).

* Note: $\rho_0 = 1$ and $\rho_k = \rho_{-k}$ for $k \neq 0$. Why?

* Existence of serial correlations implies that the return is predictable, indicating market inefficiency.

Note that

1. If r_t is strict stationary, and its first two moments are finite, then r_t is weak stationary.
2. If r_t is weak stationary and is normally distributed, then r_t is strict stationary.

Figure 23 illustrates a different example of a time-series that is not covariance stationary. In this example each value of r is weakly correlated with its previous value, at least in the first and final thirds of the dataset. In the middle third, however, each value of $\{r_t\}$ is perfectly correlated with its previous value. This implies that its variance is also not stationary, since the variance in the middle third is zero.

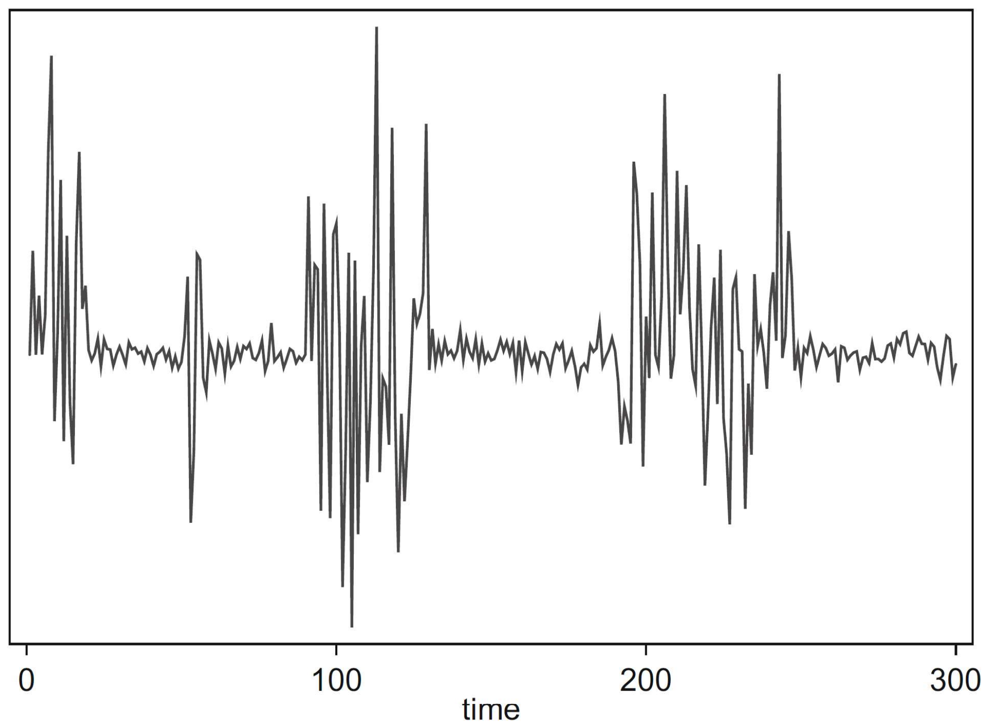


Figure 22: $\{r_t\}$ is mean stationary but not variance stationary.

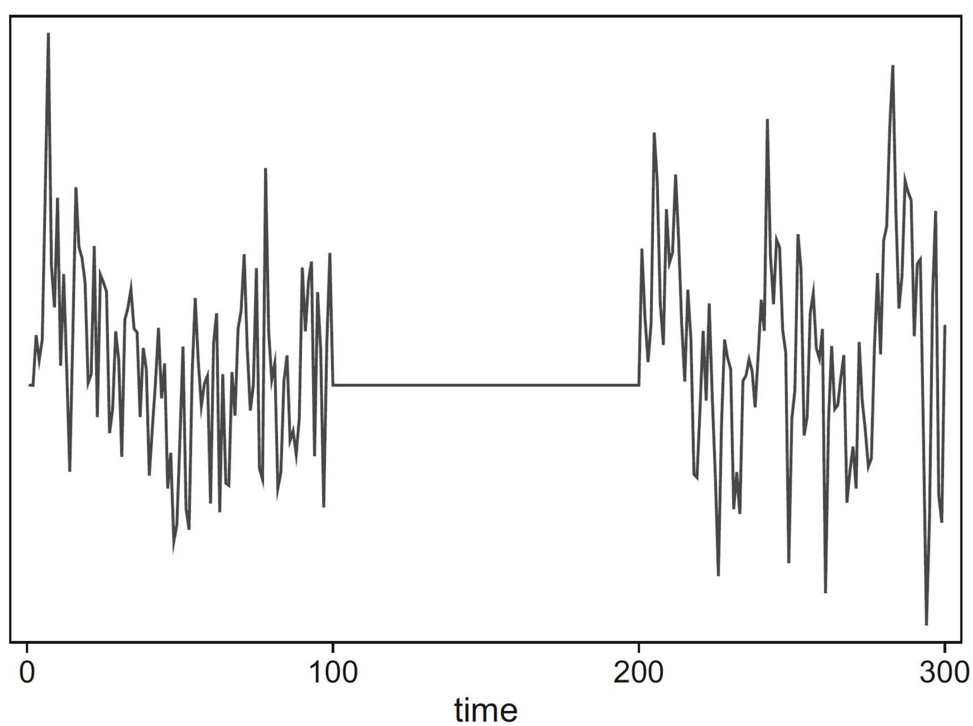


Figure 23: $\{r_t\}$ is mean stationary but neither variance nor covariance stationary.

* Sample mean and sample variance are used to estimate the mean and variance of returns:

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t, \qquad Var[r_t] = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2,$$

* Sample autocorrelation function (ACF):

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k}(r_t - \bar{r})(r_{t+k} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2},$$

where $\bar{r}$ is the sample mean and T is the sample size.

* Test zero serial correlations (market efficiency)

- Individual test: for example, $H_0 : \rho_1 = 0$ vs $H_a : \rho_1 \neq 0$. Compute

  $$t = \frac{\hat{\rho}_1}{\sqrt{1/T}}$$

  t follows a N(0,1) distribution under H_0.
  Decision rule: Reject H_0 if $|t| > Z_{\alpha/2}$, or p-value is less than α.

- Joint test (Ljung-Box statistics): $H_0 : \rho_1 = ... = \rho_m = 0$ vs $H_a : \rho_i \neq 0$ for some $i = \{1, ..., m\}$.

  $$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{T-l}$$

  $Q(m)$: Asymptotically chi-squared distribution with m degrees of freedom.
  Decision rule: Reject H_0 if $Q(m) > \mathcal{X}_m^2(\alpha)$, or p-value is less than α.

* Significant sample ACF does not necessarily imply market inefficiency. It can be because of Non-synchronous trading (ch. 5), Bid-ask bounce (ch. 5), Risk premium, etc. (ch. 3).

■ White noise: independently and identically distributed (IID) sequence a_t (with mean 0 and finite variance). 'white' refers to the way the series is distributed (i.e., independently) over time. In this case the distribution of a_t does not change over time and

$$Cov(a_t, a_{j\neq t}) = 0.$$

A white noise series is purely random. The scatter plot of a white noise series across time indicates no pattern. Therefore, forecasting the future values of a such a series is not possible. To see whether a series is not a white noise, test $H_0 : \rho_\ell = 0,\ \forall \ell > 0$. If you reject the null, it means the sries is not a white noise.

■ Suppose that every realization of a variable e is drawn independently of each other from the same identical normal distribution:

$$e_t \sim N(0, \sigma^2)$$

so that $E(e_t) = 0$, and $Var(e_t) = \sigma^2$ for all t. In this case, we say that e_t is distributed "IID Normal," or "independently and identically distributed" from a Normal distribution. If this is the case, then

$$Cov(e_t, e_{j \neq t}) = 0$$

In other words, the variable does not covary with itself across any lag; and the variable is not correlated with itself across any lag.

Example: Monthly returns of IBM stock from 1926 to 1997.

- $R_t : Q(5) = 5.4(0.37)$ and $Q(10) = 14.1(0.17)$
- $r_t : Q(5) = 5.8(0.33)$ and $Q(10) = 13.7(0.19)$

Implication: Monthly IBM stock returns do not have significant serial correlations therefore IBM returns over time can be a white noise series.

Remark: What is p-value? How to use it?

Example: Monthly returns of CRSP value-weighted index from 1926 to 1997.

- $R_t : Q(5) = 27.8$ and $Q(10) = 36.0$
- $r_t : Q(5) = 26.9$ and $Q(10) = 32.7$

All highly significant. Implication: there exist significant serial correlations in the value-weighted index returns. (Nonsynchronous trading might explain the existence of the serial correlations, among other reasons.) Similar result is also found in equal-weighted index returns.

R Demonstration: IBM monthly simple returns from 1967 to 2008.

```
> da=read.table("m-ibm6708.txt",header=T)
> ibm=da[,2]
> library(fSeries)
Loading required package: robustbase
.....
The new version of 'fSeries' has been renamed to 'timeSeries'
> acf(ibm,lag=15) <== Obtain plot of ACFs
> m1=acf(ibm,lag=15) <== Obtain plot of ACFs and more.
> names(m1)
[1] "acf" "type" "n.used" "lag" "series" "snames"
> m1$acf
[,1]
[1,] 1.0000000000
[2,] -0.0007050032
[3,] -0.0028369916
[4,] 0.0353196823
[5,] -0.0740449728
.....
[15,] 0.0013865609
[16,] -0.0360026726
> m2=pacf(ibm,lag=10) <== Obtain partial ACFs.
> names(m2)
[1] "acf" "type" "n.used" "lag" "series" "snames"
> m2$acf
[,1]
[1,] -0.0007050032
[2,] -0.0028374901
[3,] 0.0353159780
.......
[10,] 0.0304937883
```

```
> Box.test(ibm,lag=10) <=== Perform Box-Pierce test for serial correlations
Box-Pierce test
data: ibm
X-squared = 6.8338, df = 10, p-value = 0.741
> Box.test(ibm,lag=10,type='Ljung') <== Perform Box-Ljung test for serial correlations.
Box-Ljung test
data: ibm
X-squared = 6.9444, df = 10, p-value = 0.7307
```

■ Back-shift (lag) operator: A useful notation in TS analysis.

- $Br_t = r_{t-1}$ or $Lr_t = r_{t-1}$
- $B^2 r_t = r_{t-2}$ or $L^2 r_t = r_{t-2}$

Suppose that the daily log returns are

Day	1	2	3	4
r_t	0.017	0.005	0.014	0.021

Answer the following questions:

- Br_3
- $B^2 r_5$

■ A proper perspective: at a time point t

- Available data: $\{r_1, ..., r_{t-1}\} \equiv F_{t-1}$
- The return is decomposed into two parts as

  r_t = predictable part + not predictable part=function of elements of F_{t-1}+a_t

In other words, given information F_{t-1}:

$$r_t = \mu_t + a_t = E[r_t|F_{t-1}] + \sigma_t \epsilon_t$$

- μ_t: conditional mean of r_t
- a_t: shock or innovation at time t
  No one can forecast the economy certainly. We will always have shocks that are unpredictable, no matter how rich and complex our model is.
- ϵ_t: an iid sequence with mean zero and variance 1
- σ_t: conditional standard deviation (commonly called volatility in finance)

* Traditional TS modeling is concerned with μ_t.
* Model for μ_t: mean equation
* Model for σ_t^2: volatility equation

Univariate TS analysis serves two purposes

- A model for μ_t
- Understanding models for σ_t^2: properties, forecasting, etc.

- Linear time series

  Mathematically, it means r_t can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

  where μ is a constant, $\psi_0 = 1$ and $\{a_t\}$ is an iid sequence with mean zero and well-defined distribution.
  In the economic literature, a_t is the shock (or *innovation*) at time t, and $\{\psi_i\}$ are the impulse responses of r_t.

- Univariate linear time series models

  1. autoregressive (AR) models
  2. moving-average (MA) models
  3. mixed ARMA models
  4. seasonal models
  5. regression models with time series errors
  6. fractionally differenced models (long-memory)

  In this course, we use statistical methods to find models that fit the data well for making inference, e.g. prediction. On the other hand, there exists economic theory that leads to time-series models for economic variables. For instance, consider the real business-cycle theory in macroeconomics. Under some simplifying assumptions, one can show that $\ln(Y_t)$, where Y_t is the output (GDP), follows an AR(2) model. See Advanced Macroeconomics by David Romer (2006, 3rd, pp. 190).

- Important properties of a model

  - Stationarity condition
  - Basic properties: mean, variance, serial dependence
  - Empirical model building: specification, estimation, and model checking
  - Forecasting

■ **Simple AR models:** (Regression with lagged variables.)

Motivating example: Quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991. The model that fits the data (discussed later) is

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.01,$$

This is called an AR(3) model because the growth rate r_t depends on the growth rates of the past **three** quarters. How do we specify this model from the data? Is it adequate for the data? What are the implications of the model? Statistical significance vs economic significance. These are the questions we shall address in this lecture.

AR(1) model: The fact that the monthly returns r_t of CRSP value-weighted index has a statistically significant lag-1 autocorrelation indicates that the lagged return r_{t-1} might be useful in predicting r_t.

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, where $\{a_t\}$ is assumed to be white noise with mean 0 and variance σ_a^2, ϕ_0 and ϕ_1 are real numbers, which are referred to as "parameters" (to be estimated from the data). For example,

$$r_t = 0.009 + 0.107r_{t-1} + a_t, \quad \hat{\sigma}_a^2 = 0.003$$

2. Conditional mean: $E[r_t|r_{t-1}] = \phi_0 + \phi_1 r_{t-1}$

3. Conditional variance: $\mathrm{Var}[r_t|r_{t-1}] = \sigma_a^2$

4. (Unconditional) Mean: $E[r_t] = \frac{\phi_0}{1-\phi_1}$

5. (Unconditional) Variance: $Var[r_t] = \frac{\sigma_a^2}{1-\phi_1^2}$

6. Stationarity: Not all AR processes are stationary. Some grow without limit. Some have variances which change over time. Necessary and sufficient condition for AR(1) is $|\phi_1| < 1$. Why?

7. Alternative representation: Let $E[r_t] = \mu$ be the mean of r_t so that $\mu = \frac{\phi_0}{1-\phi_1}$. Equivalently, $\phi_0 = \mu(1 - \phi_1)$. Plugging in the model, we have

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + a_t$$

   This model also has two parameters (μ and ϕ_1). It explicitly uses the mean of the series. It is less commonly used in the literature, but is the model representation used in R.

8. Autocorrelations: $\rho_1 = \phi_1$, $\rho_2 = \phi_1^2$, etc. In general

$$\rho_k = \phi_1^k,$$

   and ACF decays exponentially as k increases.

9. A compact form: $(1-\phi_1 B)r_t = \phi_0 + a_t$.

10. Forecasts: Given that we have estimated an AR(1) model, how can we use it to forecast future values of r_t. Suppose the forecast origin is n. For simplicity, we shall use the model representation in (1) and write $x_t = r_t - \mu$. The model then becomes $x_t = \phi_1 x_{t-1} + a_t$. Note that forecast of r_t is simply the forecast of x_t plus μ.

    – 1-step ahead forecast at time n:

    $$\hat{x}_n(1) = \phi_1 x_n$$

    – 1-step ahead forecast error:

    $$\hat{e}_n(1) = x_{n+1} - \hat{x}_n(1) = a_{n+1}$$

    Thus, a_{n+1} is the un-predictable part of x_{n+1}. It is the shock at time $n+1$!

    – Variance of 1-step ahead forecast error:

    $$Var[e_n(1)] = Var[a_{n+1}] = \sigma_a^2$$

    – 2-step ahead forecast:

    $$\hat{x}_n(2) = \phi_1 \hat{x}_n(1) = \phi_1^2 x_n$$

    – 2-step ahead forecast error:

    $$\hat{e}_n(2) = x_{n+2} - \hat{x}_n(2) = a_{n+2} + \phi_1 a_{n+1}$$

    – Variance of 2-step ahead forecast error:

    $$Var[e_n(2)] = Var[a_{n+2} + \phi_1 a_{n+1}] = (1+\phi_1^2)\sigma_a^2$$

    which is greater than or equal to $Var[e_n(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

    – Behavior of multi-step-ahead forecasts: In general, for the l-step ahead forecast at origin n, we have

    $$\hat{x}_n(l) = \phi_1^l x_n,$$

    the forecast error

    $$\hat{e}_n(l) = a_{n+l} + \phi_1 a_{n+l-1} + ... + \phi_1^{l-1} a_{n+1}$$

    and the variance of forecast error

    $$Var[e_n(l)] = (1 + \phi_1^2 + ... + \phi_1^{2(l-1)})\sigma_a^2$$

    In particular, as $l \longrightarrow \infty$

    $$\hat{x}_n(l) \longrightarrow 0, \quad i.e., \quad \hat{r}_n(l) \longrightarrow \mu$$

    This is called the mean-reversion of the AR(1) process. The variance of forecast error approaches

    $$Var[e_n(l)] = \frac{1}{1-\phi_1^2}\sigma_a^2 = Var[r_t]$$

    In practice, it means that for the long-term forecasts serial dependence is not important. The forecast is just the sample mean and the uncertainty is simply the uncertainty about the series.

AR(2) model:

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$, or
$$(1 - \phi_1 B - \phi_2 B^2) r_t = \phi_0 + a_t$$

2. Mean: $E[r_t] = \frac{\phi_0}{1-\phi_1-\phi_2}$

3. Mean-adjusted format: Using $\phi_0 = \mu - \phi_1\mu - \phi_2\mu$, we can write the AR(2) model as
$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t$$
This form is often used in the finance literature to highlight the mean-reverting property of a stationary AR(2) model.

4. Autocorrelations: $\rho_0 = 1$, $\rho_1 = \frac{\phi_1}{1-\phi_2}$,
$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 2.$$

5. Stationarity condition:(factor of polynomial)

6. Characteristic equation: $(1 - \phi_1 x - \phi_2 x^2) = 0$

7. Stochastic business cycle: if $\phi_1^2 + 4\phi_2 < 0$, then r_t shows characteristics of business cycles with average length
$$k = \frac{2\pi}{cos^{-1}(\phi_1/2\sqrt{-\phi_2})}$$
where the cosine inverse is stated in radian. If we denote the solutions of the polynomial as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$ and $\phi_2 = -(a^2+b^2)$ so that
$$k = \frac{2\pi}{cos^{-1}(a/\sqrt{a^2+b^2})}$$

8. Forecasts: Similar to AR(1) models. R can automate these calculations for us (After estimating the model, use R predict command *predict*).

Discussion: (Reference only)

An AR(2) model can be written as an AR(1) model if one expands the dimension. Specifically, we have

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t \tag{3.1}$$
$$r_{t-1} - \mu = r_{t-1} - \mu \tag{3.2}$$

Now, putting the two equations together, we have

$$\begin{bmatrix} r_t - \mu \\ r_{t-1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} - \mu \\ r_{t-2} - \mu \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}. \tag{3.3}$$

This is a 2-dimensional AR(1) model. Several properties of the AR(2) model can be obtained from the expanded AR(1) model.

■ **Building an AR model:**

- Order specification

  1. Partial ACF: (naive, but effective)

     * Use consecutive fittings (See Text (Section 2.4.2) for details). Theoretical PACFs are similar to ACFs, except they remove the effects of other lags. That is, the PACF at lag 2 filters out the effect of autocorrelation from lag 1. Likewise, the partial autocorrelation at lag 3 filters out the effect of autocorrelation at lags 2 and 1.
     * Key feature: PACF cuts off at lag p for an AR(p) model. For an AR(p), the PACF has spikes at lags 1 through p, and then zeros at lags greater than p.
     * Illustration: See the PACF of the U.S. quarterly growth rate of GNP.

     We can make use of various "information criteria". In a nutshell, information criteria penalize the log-likelihood by various amounts, depending on the number of observations and the number of estimated statistics:

  2. Akaike information criterion $AIC(\ell) = -\frac{2}{T}\ln(\text{likelihood}) + \frac{2}{T} \times ($ number of parameters). For a Gaussian AR(ℓ) model, AIC reduces to

     $$AIC(\ell) = \ln(\sigma_\ell^2) + \frac{2\ell}{T}$$

     for an AR(ℓ) model, where σ_ℓ^2 is the MLE of residual variance.
     Find the AR order with minimum AIC for $\ell \in [0, ..., P]$.

  3. BIC criterion:

     $$BIC(\ell) = \ln(\sigma_\ell^2) + \frac{\ell \ln(T)}{T}$$

- Needs a constant term? Check the sample mean.
- Estimation: least squares method or maximum likelihood method
- Model checking:

  1. calculate residuals: observations minus the fit, i.e. 1-step ahead forecast errors at each time point.
  2. Residual should be close to white noise if the model is adequate. Use Ljung-Box statistics of residuals, but degrees of freedom is m-g, where g is the number of AR coefficients used in the model.

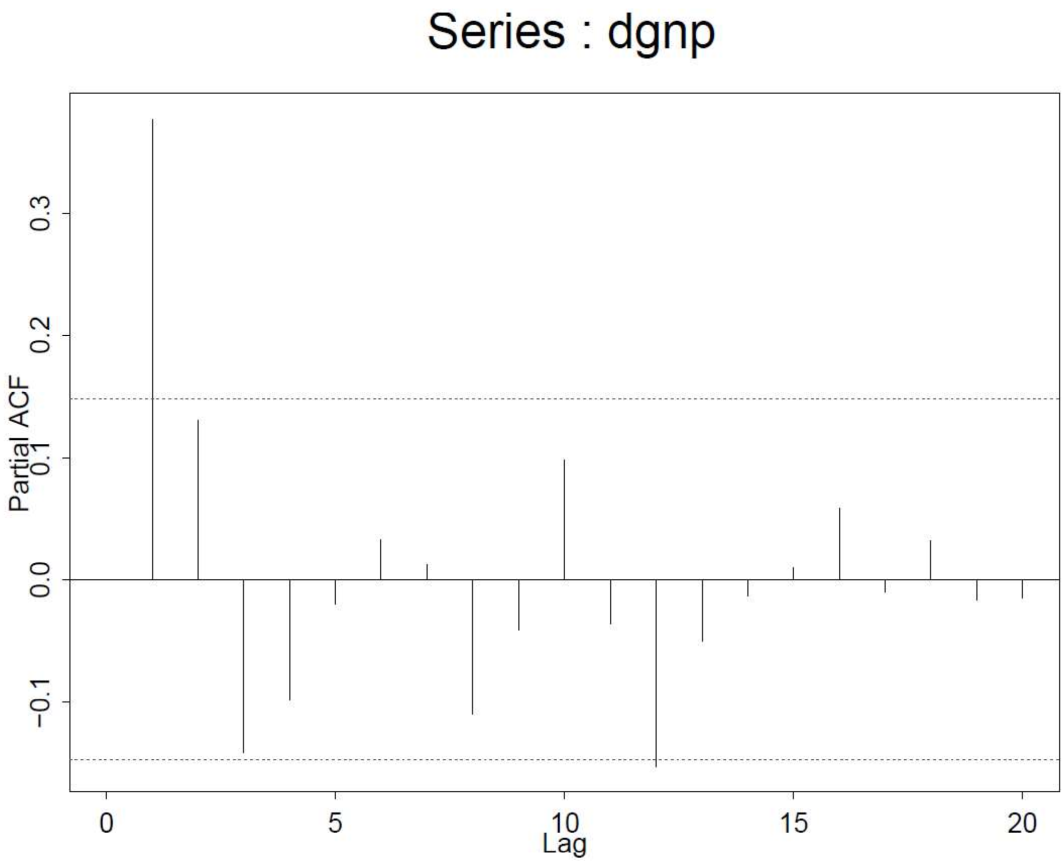
     Note: The Box-Jenkins procedure is concluded by verifying that the estimated residuals are white noise. This implies that there is no leftover structure to the data that we have neglected to model. If the residuals are not white noise, then Box and Jenkins recommend modifying the model, re-estimating, and re-examining the residuals.

  3. Goodness of fit: R^2 and adjusted R^2. A commonly used statistic to measure goodness of fit of a stationary model is the R square (R^2) defined as

     $$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}}$$

Typically, a larger R^2 indicates that the model provides a closer fit to the data. However, this is only true for a stationary time series. For the unit-root non-stationary series discussed later, R^2 of an AR(1) fit converges to one when the sample size increases to infinity, regardless of the true underlying model of r_t.

– Example: Analysis of U.S. GNP growth rate series.



R Demonstration: U.S. GNP growth rate series

```
> setwd("C:/Users/rst/teaching/bs41202/sp2015")
> library(fBasics)
> da=read.table("dgnp82.dat")
> x=da[,1]
> par(mfcol=c(2,2))  put 4 plots on a page
> plot(x,type='l') first plot
> plot(x[1:175],x[2:176])  2nd plot
> plot(x[1:174],x[3:176])  3rd plot
> acf(x,lag=12)% 4th plot
> pacf(x,lag.max=12)  Compute PACF (not shown in this handout)
> Box.test(x,lag=10,type='Ljung')  Compute Q(10) statistics
Box-Ljung test
data: x
X-squared = 43.2345, df = 10, p-value = 4.515e-06


> m1=ar(x,method='mle')  Automatic AR fitting using AIC criterion.
> m1
Call: ar(x = x, method = "mle")
Coefficients:
1 2 3  An AR(3) is specified.
0.3480 0.1793 -0.1423
```

```
Order selected 3 sigma^2 estimated as 9.427e-05
> names(m1)
[1] "order" "ar" "var.pred" "x.mean" "aic"
[6] "n.used" "order.max" "partialacf" "resid" "method"
[11] "series" "frequency" "call" "asy.var.coef"
> plot(m1$resid,type='l')  Plot residuals of the fitted model (not shown)
> Box.test(m1$resid,lag=10,type='Ljung')  Model checking
Box-Ljung test
data: m1$resid
X-squared = 7.0808, df = 10, p-value = 0.7178
> m2=arima(x,order=c(3,0,0))  %Another approach with order given.
> m2
Call:
arima(x = x, order = c(3, 0, 0))
Coefficients:
ar1 ar2 ar3 intercept  % Fitted model is y(t)=0.348y(t-1)+0.179y(t-2) -0.142y(t-3)+a(t), where y(t) = x(t)-0.0077
0.3480 0.1793 -0.1423 0.0077
s.e. 0.0745 0.0778 0.0745 0.0012

sigma^2 estimated as 9.427e-05: log likelihood = 565.84, aic = -1121.68
> names(m2)
[1] "coef" "sigma2" "var.coef" "mask" "loglik" "aic"
[7] "arma" "residuals" "call" "series" "code" "n.cond"
[13] "model"
> Box.test(m2$residuals,lag=10,type='Ljung')
Box-Ljung test
data: m2$residuals
X-squared = 7.0169, df = 10, p-value = 0.7239
> plot(m2$residuals,type='l')  % Residual plot
> tsdiag(m2)  % obtain 3 plots of model checking (not shown in handout).
> p1=c(1,-m2$coef[1:3])  % Further analysis of the fitted model.
> roots=polyroot(p1)
> roots
[1] 1.590253+1.063882e+00i -1.920152-3.530887e-17i 1.590253-1.063882e+00i
> Mod(roots)
[1] 1.913308 1.920152 1.913308
> k=2*pi/acos(1.590253/1.913308)
> k
[1] 10.65638
> predict(m2,8)  %Prediction 1-step to 8-step ahead.
$pred
Time Series:
Start = 177
End = 184
Frequency = 1
[1] 0.001236254 0.004555519 0.007454906 0.007958518
[5] 0.008181442 0.007936845 0.007820046 0.007703826
$se
Time Series:
Start = 177
End = 184
Frequency = 1
[1] 0.009709322 0.010280510 0.010686305 0.010688994
[5] 0.010689733 0.010694771 0.010695511 0.010696190
$
```

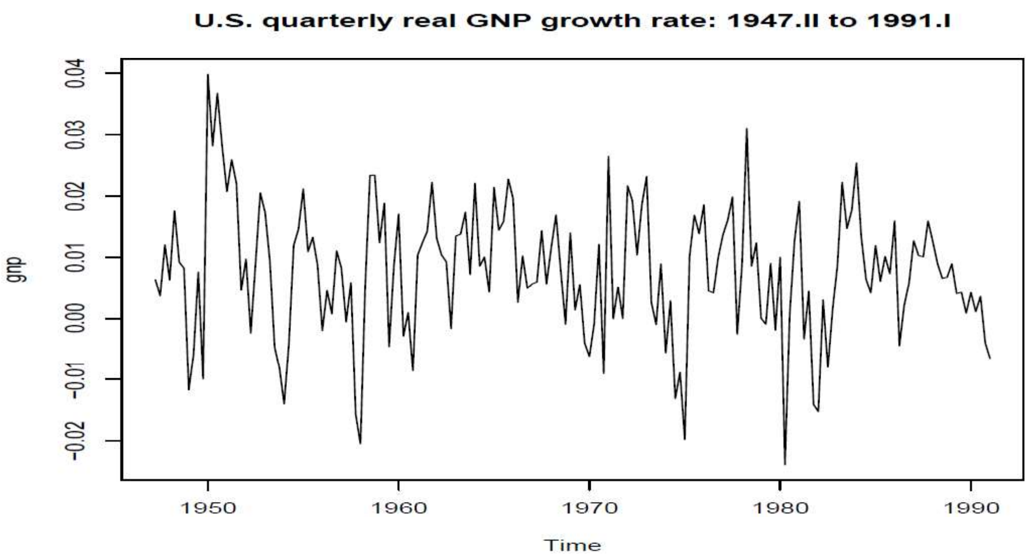


Figure 24: U.S. quarterly growth rate of real GNP: 1947-1991

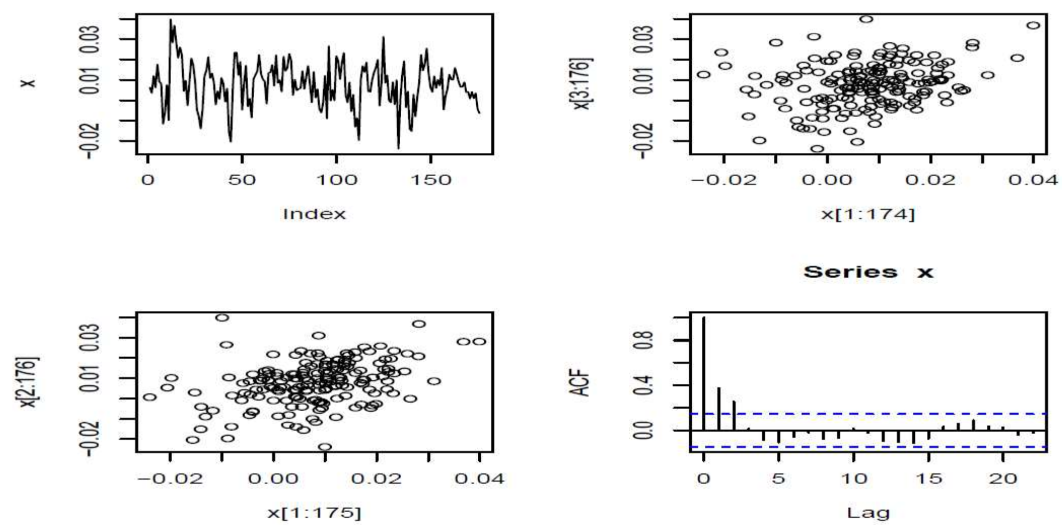


Figure 25: Various plots of U.S. quarterly growth rate of real GNP: 1947-1991

Another example: Monthly U.S. unemployment rate from January 1948 to February 2015. I use this example to emphasize two messages: (1) Simple model is often preferred; (2) seasonal adjustment does not necessarily remove all seasonality.

R Demonstration:

```
require(quantmod)
> get Symbols("UNRATE",src="FRED")
> chartSeries(UNRATE)
> rate <- as.numeric(UNRATE[,1]) ## Use a regular vector, instead of an ‘‘xts’’ object
> ts.plot(rate)
> acf(rate)
> zt=diff(rate)
> acf(zt)
> pacf(zt)
> m1=ar(zt,lag.max=20,method="mle")
> names(m1)
[1] "order" "ar" "var.pred" "x.mean" "aic"
[6] "n.used" "order.max" "partialacf" "resid" "method"
[11] "series" "frequency" "call" "asy.var.coef"
> m1$order
[1] 12
> t.test(zt) ### Checking for mean value
One Sample t-test
data: zt
t = 0.3481, df = 804, p-value = 0.7278
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.01210074 0.01731813
sample estimates:
mean of x
0.002608696
> m2=arima(rate,order=c(12,1,0))
> m2
Call:
arima(x = rate, order = c(12, 1, 0))
Coefficients:
ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9
0.0181 0.2163 0.1484 0.0964 0.1352 0.0006 -0.0341 0.0150 0.0045
s.e. 0.0350 0.0351 0.0357 0.0363 0.0364 0.0368 0.0367 0.0365 0.0364
ar10 ar11 ar12
-0.0893 0.0299 -0.1327
s.e. 0.0359 0.0353 0.0353
sigma^2 estimated as 0.03744: log likelihood = 179.62, aic = -333.25
> tsdiag(m2,gof=24)
> c1=c(0,NA,NA,NA,NA,0,0,0,0,NA,0,NA) ### Remove insignificant parameters
> m2a=arima(rate,order=c(12,1,0),fixed=c1)
Warning message:
In arima(rate, order = c(12, 1, 0), fixed = c1) :
some AR parameters were fixed: setting transform.pars = FALSE
> m2a
Call:
arima(x = rate, order = c(12, 1, 0), fixed = c1)
Coefficients:
ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8 ar9 ar10 ar11
0 0.2130 0.1539 0.0988 0.1340 0 0 0 0 -0.0876 0
s.e. 0 0.0347 0.0339 0.0341 0.0352 0 0 0 0 0.0346 0
ar12
-0.1328
s.e. 0.0342
sigma^2 estimated as 0.03754: log likelihood = 178.58, aic = -343.15
> predict(m2a,4)
$pred
Time Series:
Start = 807
End = 810
Frequency = 1
[1] 5.478118 5.515175 5.440772 5.474257
$se
Time Series:
Start = 807
End = 810
```

```
Frequency = 1
[1] 0.1937406 0.2739905 0.3609757 0.4477078
> m3=arima(rate,order=c(2,1,1),seasonal=list(order=c(1,0,1),period=12))
> m3
Call:
arima(x = rate, order = c(2, 1, 1), seasonal = list(order = c(1, 0, 1), period = 12))
Coefficients:
ar1 ar2 ma1 sar1 sma1
0.6046 0.2313 -0.6016 0.5453 -0.8146
s.e. 0.0603 0.0384 0.0554 0.0676 0.0488
sigma^2 estimated as 0.03587: log likelihood = 195.51, aic = -379.02
> tsdiag(m3,gof=24)
> predict(m3,4)
$pred
Time Series:
Start = 807
End = 810
Frequency = 1
[1] 5.503355 5.475984 5.416243 5.386766
$se
Time Series:
Start = 807
End = 810
Frequency = 1
[1] 0.1893917 0.2682392 0.3560308 0.4414087

$
```

Simulation in R:

```
y1=arima.sim(model=list(ar=c(1.3,-.4)),1000)
y2=arima.sim(model=list(ar=c(.8,-.7)),1000)
```

Check the ACF and PACF of the above two simulated series.

■ **Moving-average (MA) Model:**

Model with finite memory!

AR models had autocorrelated r's because current r depended directly upon lagged values of r. MA models, on the other hand have autocorrelated r's because the errors are, themselves, autocorrelated.

Some daily stock returns have minor serial correlations and can be modeled as MA or AR models.

MA(1) Model: The simplest type of MA model is

- Form: $r_t = \mu + a_t - \theta a_{t-1},\ a_t \sim iidN(0, \sigma_a^2)$
  It will be useful to differentiate between the errors $(a_t - \theta a_{t-1})$ from the random shocks (a_t). The error terms are autocorrelated. The shocks (a_t) are presumed to be white noise. The important thing to note is that the action in this model lies in the fact that the errors have a direct effect on or beyond the immediate term.
- Stationarity: always stationary
- Mean (or expectation): $E[r_t] = \mu$
- Variance: $Var[r_t] = (1 + \theta^2)\sigma_a^2$
- Autocovariance:
  1. Lag 1: $Cov(r_t, r_{t-1}) = -\theta\sigma_a^2$
  2. Lag ℓ: $Cov(r_t, r_{t-\ell}) = 0$, for $\ell > 1$.
     Thus, r_t is not related to r_{t-2}, r_{t-3}, ...
- ACF: $\rho_1 = \frac{-\theta}{1+\theta^2},\ \rho_\ell = 0$, for $\ell > 1$.
  Finite memory! MA(1) models do not remember what happen two time periods ago.
- Forecast (at origin $t = n$):
  1. 1-step ahead: $\hat{r}_n(1) = \mu - \theta a_n$. Why? Because at time n, a_n is known, but a_{n+1} is not.
  2. 1-step ahead forecast error: $e_n(1) = a_{n+1}$ with variance σ_a^2.
  3. Multi-step ahead: $\hat{r}_n(\ell) = \mu$, for $\ell > 1$.
     Thus, for an MA(1) model, the multi-step ahead forecasts are just the mean of the series. Why? Because the model has memory of 1 time period.
  4. Multi-step ahead forecast error:
     $$e_n(\ell) = a_{n+\ell} - \theta a_{n+\ell-1}$$
  5. Variance of multi-step ahead forecast error:
     $$(1 + \theta^2)\sigma_a^2 = \text{variance of } r_t$$
- Invertibility:
  * Concept: r_t is a proper linear combination of a_t and the past observations $\{r_{t-1}, r_{t-2}, ...\}$.

* Why is it important? It provides a simple way to obtain the shock a_t. For an invertible model, the dependence of r_t on $r_{t-\ell}$ converges to zero as ℓ increases.
* Condition: $|\theta| < 1$
* Invertibility of MA models is the dual property of stationarity for AR models.

MA(2) Model:

- Form: $r_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$. or

$$r_t = \mu + (1 - \theta_1 B - \theta_2 B^2) a_t$$

- Stationary with $E(r_t) = \mu$
- Variance: $Var[r_t] = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$
- ACF: $\rho_2 \neq 0$, but$\rho_\ell = 0$, for $\ell > 2$.
- Forecasts go to the mean after 2 periods.

Building an MA model

- Specification: Use sample ACF

  Sample ACFs are all small after lag q for an MA(q) series. (See test of ACF.)
- Constant term? Check the sample mean.
- Estimation: use maximum likelihood method
  * Conditional: Assume $a_t = 0$ for $t \leq 0$.
  * Exact: Treat a_t with $t \leq 0$ as parameters, estimate them to obtain the likelihood function.

    Exact method is preferred, but it is more computing intensive.
- Model checking: examine residuals (to be white noise)
- Forecast: use the residuals as $\{a_t\}$ (which can be obtained from the data and fitted parameters) to perform forecasts.

Model form in R: R parameterizes the MA(q) model as

$$r_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}$$

instead of the usual minus sign in θ. Consequently, care needs to be exercised in writing down a fitted MA parameter in R. For instance, an estimate $\hat{\theta}_1 = 0.5$ of an MA(1) in R indicates the model is $r_t = a_t + 0.5a_{t-1}$.

Example:Daily log return of the value-weighted index

R Demonstration:

```
> setwd("C:/Users/rst/teaching/bs41202/sp2015")
> library(fBasics)
> da=read.table("d-ibmvwew6202.txt")
> dim(da)
[1] 10194 4
> vw=log(1+da[,3])*100 % Compute percentage log returns of the vw index.
> acf(vw,lag.max=10) % What type of process might have generated this data?
> m1=arima(vw,order=c(0,0,1)) % fits an MA(1) model
> m1
Call:
arima(x = vw, order = c(0, 0, 1))
Coefficients:
ma1 intercept
0.1465 0.0396 % The model is vw(t) = 0.0396+a(t)+0.1465a(t-1).
s.e. 0.0099 0.0100
sigma^2 estimated as 0.7785: log likelihood = -13188.48, aic = 26382.96
> tsdiag(m1)
> predict(m1,5)
$pred
Time Series:
Start = 10195
End = 10199
Frequency = 1
[1] 0.05036298 0.03960887 0.03960887 0.03960887 0.03960887
$se
Time Series:
Start = 10195
End = 10199
Frequency = 1
[1] 0.8823290 0.8917523 0.8917523 0.8917523 0.8917523
```

Summary: ACFs and PACFs each come in two flavors: theoretical and empirical. The former is implied by a model; the latter is a characteristic of the data. We can compare (a) the empirical ACFs and PACFs that we estimate directly from data without using a model, with (b) the theoretical ACFs and PACFs that are associated with a particular model. Then, we only need to see how they match up. That is, we can be fairly certain that the data were generated from a particular type of process (model) if the empirical ACF matches up with that of a particular model's theoretical ACFs.

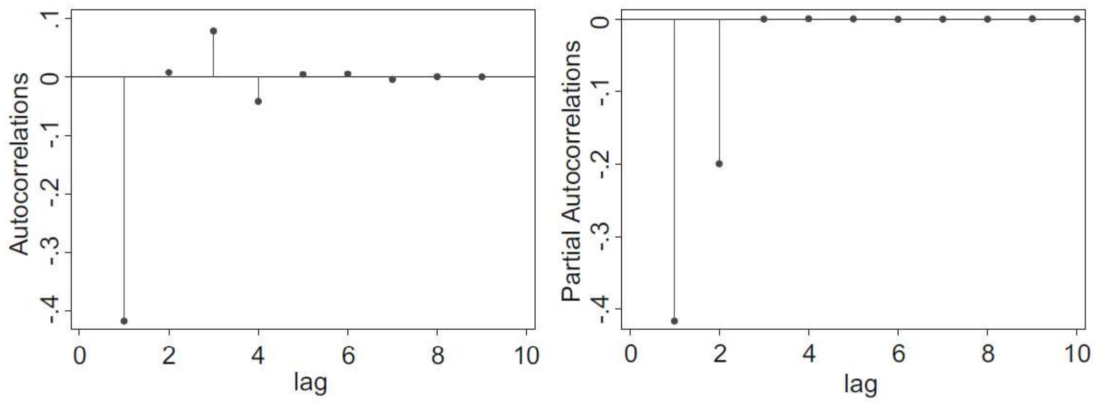
1. In MA models, ACF is useful in specifying the order of the model, q, because ACF cuts off at lag q for a MA(q) series.

2. In AR models, PACF is useful in specifying the order of the model, p, because PACF cuts off at lag p for an AR(p) series.

3. MA series are always weak stationary, but for an AR model to ba stationary, all its characteristic roots must be greater than 1 in modulus.

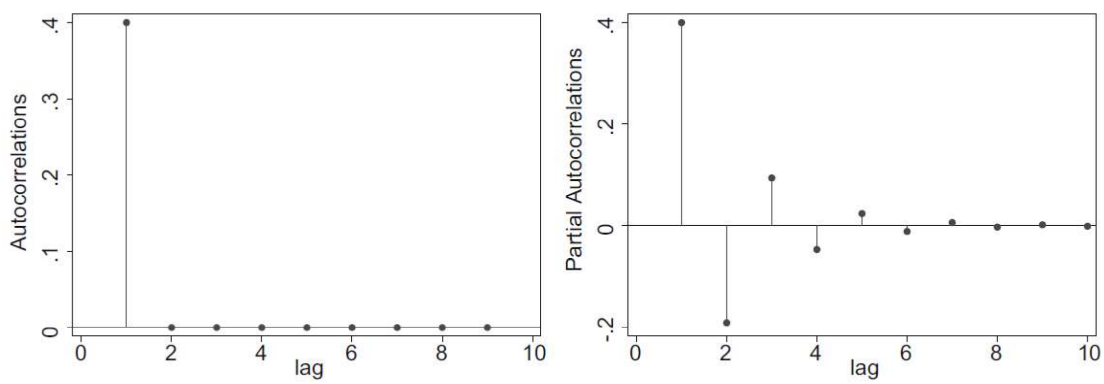
4. In MA models, to have an invertible model, we must have $|\theta_i| < 1, \quad \forall i$.

5. For stationary AR and MA models, the multi-step forecasts converge to the unconditional mean of the series (mean reversion), and the variance of the forecasts errors converges to the unconditional variance of the series as the forecast horizon increases ($\ell \to \infty$).



(a) (b)

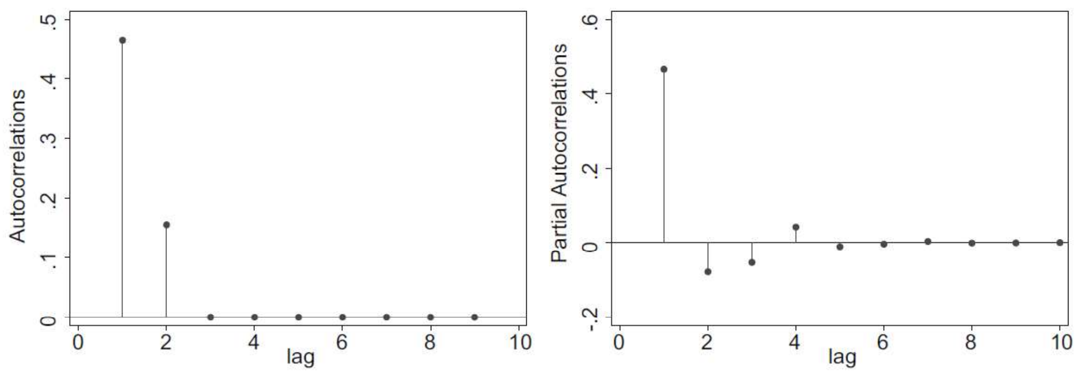
Theoretical (**a**) ACF and (**b**) PACF of AR(2): $X_t = -0.50X_{t-1} - 0.20X_{t-2} + e_t$



(a) (b)

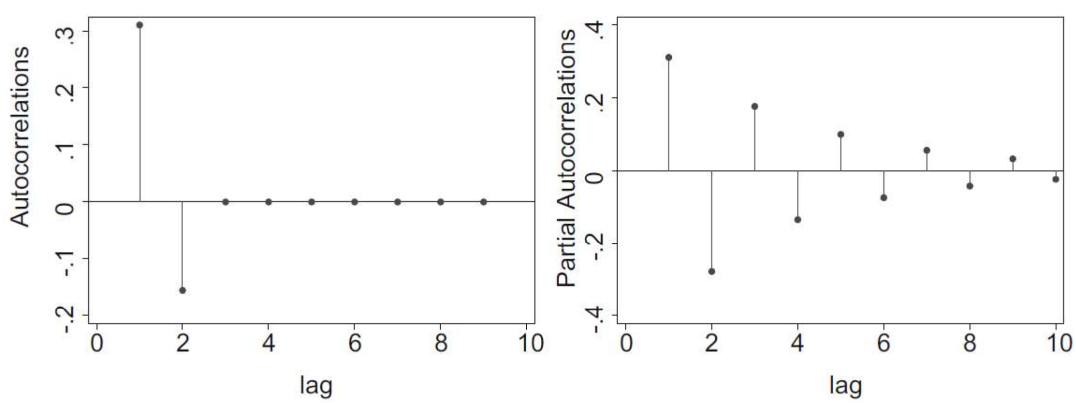
Theoretical (**a**) ACF and (**b**) PACF of MA(1): $X_t = u_t + 0.50u_{t-1}$

Figure 26: ACF and PACF of MA and AR models.



(a) (b)

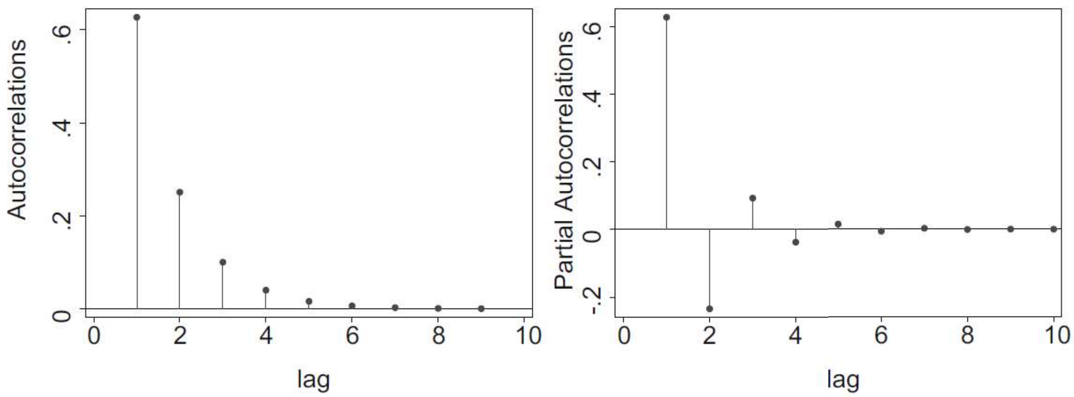
Theoretical (**a**) ACF and (**b**) PACF of MA(2): $X_t = u_t + 0.50u_{t-1} + 0.20u_{t-2}$



(a) (b)

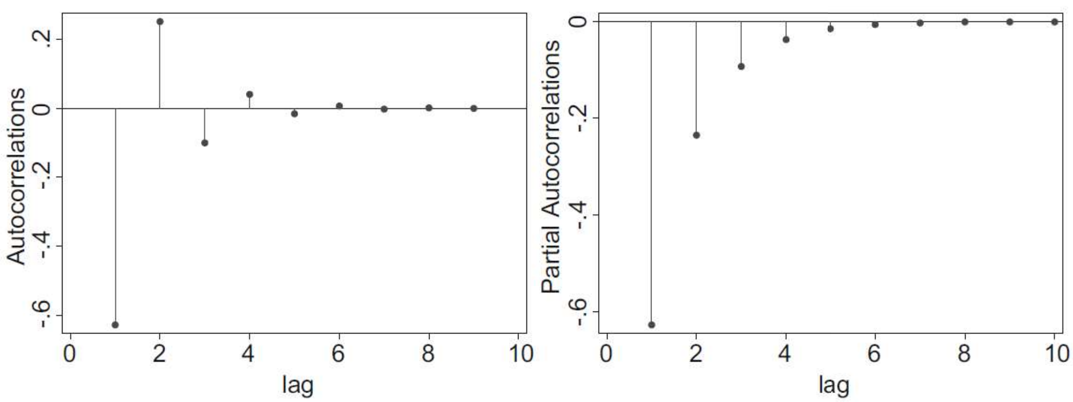
Theoretical (**a**) ACF and (**b**) PACF of MA(2): $X_t = u_t + 0.50u_{t-1} - 0.20u_{t-2}$

Figure 27: ACF and PACF of MA models with different parameters.



(a) (b)

Theoretical (**a**) ACF and (**b**) PACF of $X_t = 0.40X_{t-1} + u_t + 0.40u_{t-1}$



(a) (b)

Theoretical (**a**) ACF and (**b**) PACF of $X_t = -0.40X_{t-1} + u_t - 0.40u_{t-1}$

Figure 28: We need a more complicated model than MA and AR. ACF does not cut off to 0 at any lag, nor does PACF.