Seasonal Time Series: TS with periodic patterns and useful in

• predicting quarterly earnings
• pricing weather-related derivatives
• analysis of transactions data (high-frequency data), e.g., U-shaped pattern in intraday data

Example 1. Monthly U.S. Housing Starts from January 1992 to December 2006. The data are in logarithm. See Figure 1 and compute the sample ACF of the series and its differenced data.

Example 2. Quarterly earnings of Johnson & Johnson
See the time plot, Figures 2 and 3, and sample ACFs

Example 3. Quarterly earning per share of FedEx from the fourth quarter of 1991 to the fourth quarter of 2006.

Multiplicative model: Consider the housing-starts series. Let $y_t$ be the monthly data. Denoting 1992 as year 0, we can write the time index as $t = \text{year} + \text{month}$, e.g., $y_1 = y_{0,1}$, $y_2 = y_{0,2}$, and $y_{14} = y_{1,2}$, etc. The multiplicative model is based on the following consideration:
Figure 1: Time plot of monthly US single-family housing starts: 1992-2006. Data obtained from US Bureau of the Census.

Figure 2: Time plot of quarterly earnings of Johnson and Johnson: 1960-1980
Figure 3: Time plot of quarterly logged earnings of Johnson and Johnson: 1960-1980

Figure 4: Time plot of quarterly earnings per share of FedEx: 1991.IV to 2006.IV
The column dependence is the usual lag-1, lag-2, ... dependence. That is, monthly dependence. We call them the regular dependence. The row dependence is the year-to-year dependence. We call them the seasonal dependence.

*Multiplicative* model says that the regular and seasonal dependence are orthogonal to each other.

**Airline model** (for quarterly series)

- **Form:**
  \[ r_t - r_{t-1} - r_{t-4} + r_{t-5} = a_t - \theta_1 a_{t-1} - \theta_4 a_{t-4} + \theta_1 \theta_4 a_{t-5} \]

  or

  \[ (1 - B)(1 - B^4)r_t = (1 - \theta_1 B)(1 - \theta_4 B^4)a_t \]

- **Define the differenced series** \( w_t \) as

  \[ w_t = r_t - r_{t-1} - r_{t-4} + r_{t-5} = (r_t - r_{t-1}) - (r_{t-4} - r_{t-5}). \]

  It is called *regular* and *seasonal* differenced series.
Figure 5: Forecast plot for the quarterly earnings of Johnson and Johnson. Data: 1960-1980, Forecasts: 1981-82.

- ACF of $w_t$ has a nice symmetric structure (see the text), i.e. $\rho_{s-1} = \rho_{s+1} = \rho_1 \rho_s$. Also, $\rho_\ell = 0$ for $\ell > s + 1$.

- This model is widely applicable to many many seasonal time series.

- Multiplicative model means that the regular and seasonal dependences are roughly orthogonal to each other.

- Forecasts: exhibit the same pattern as the observed series. See Figure 5.

**Example** Detailed analysis of J&J earnings.

**R Demonstration:** output edited.

```r
> setwd("C:/Users/rst/teaching/bs41202/sp2011")
```
> x = ts(scan("q-earn-jnj.txt"), frequency = 4, start = c(1960, 1)) % create a time series object.
> plot(x) % Plot data with calendar time
> y = log(x) % Natural log transformation
> plot(y) % plot data
> c1 = paste(c(1:4))
> points(y, pch = c1) % put circles on data points.
> par(mfcol = c(2, 1)) % two plots per page
> acf(y, lag.max = 16)
> y1 = as.vector(y) % Creates a sequence of data in R
> acf(y1, lag.max = 16)
> dy1 = diff(y1) % regular difference
> acf(dy1, lag.max = 16)
> sdy1 = diff(dy1, 4) % seasonal difference
> acf(sdy1, lag.max = 12)

> m1 = arima(y1, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 4)) % Airline
% model in R.
> m1
Call:
arima(x = y1, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 4))
Coefficients:
ma1  sma1
-0.6809  -0.3146  % The fitted model is (1-B^4)(1-B)R(t) =
s.e.  0.0982  0.1070  % (1-0.68B)(1-0.31B^4)a(t), var[a(t)] = 0.00793.
sigma^2 estimated as 0.00793: log likelihood = 78.38, aic = -150.75
> par(mfcol = c(1, 1)) % One plot per page
> tsdiag(m1) % Model checking
> f1 = predict(m1, 8) % prediction
> names(f1)
[1] "pred" "se"
> f1
$pred  % Point forecasts
Time Series:
Start = 85
End = 92
Frequency = 1

$se  % standard errors of point forecasts
Time Series:
Start = 85
End = 92
Frequency = 1
[1] 0.08905414 0.09347895 0.09770358 0.10175295 0.13548765 0.14370550
You can use ‘‘foreplot’’ to obtain plot of forecasts.

Consider monthly series with period 12. Airline model becomes

\[(1 - B)(1 - B^{12})r_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})a_t.\]

What is the pattern of ACF?

**Regression Models with Time Series Errors**

- Has many applications

- Impact of serial correlations in regression is often overlooked. It may introduce biases in estimates and in standard errors, resulting in unreliable t-ratios.

- Detecting residual serial correlation: Use Q-stat instead of DW-statistic, which is not sufficient!

- Joint estimation of all parameters is preferred.

- Proper analysis: see the illustration below.

**Example.** U.S. weekly interest rate data: 1-year and 3-year constant maturity rates. Data are shown in Figure 6.

**R Demonstration:** output edited.

```r
> library(timeSeries)
> setwd("C:/Users/rst/teaching/bs41202/sp2011")
> da=read.table("w-gs1n36299.txt") % load the data
> r1=da[,1] % 1-year rate
> r3=da[,2] % 3-year rate
```
Figure 6: Time plots of U.S. weekly interest rates: 1-year constant maturity rate (solid line) and 3-year rate (dashed line).

```r
> plot(r1,type='l') % Plot the data
> lines(1:1967,r3,lty=2)
> plot(r1,r3) % scatter plot of the two series

> m1=lm(r3~r1) % Fit a regression model with likelihood method.
> summary(m1)
Call:
  lm(formula = r3 ~ r1)
Residuals:
            Min       1Q   Median       3Q      Max
-1.812147  -0.402280   0.003097  0.402588  1.338746
Coefficients:  
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.910687   0.032250  28.243  < 2e-16 ***
r1                   0.923854   0.004389 210.512  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.538 on 1965 degrees of freedom
Multiple R-Squared: 0.9575, Adjusted R-squared: 0.9575
```
F-statistic: 4.431e+04 on 1 and 1965 DF, p-value: < 2.2e-16

> acf(m1$residuals)
> c3=diff(r3)
> c1=diff(r1)
> plot(c1,c3)

> m2=lm(c3~c1) % Fit a regression with likelihood method.
> summary(m2)
Call:
  lm(formula = c3 ~ c1)
Residuals:
   Min     1Q Median     3Q    Max
-0.3806040 -0.0333840 -0.0005428 0.0343681 0.4741822
Coefficients:            Estimate Std. Error t value Pr(>|t|)
  (Intercept) 0.0002475  0.0015380   0.161  0.872     
c1   0.7810590  0.0074651 104.628 <2e-16 ***
---
Residual standard error: 0.06819 on 1964 degrees of freedom
Multiple R-Squared: 0.8479,    Adjusted R-squared: 0.8478
F-statistic: 1.095e+04 on 1 and 1964 DF, p-value: < 2.2e-16

> acf(m2$residuals)
> plot(m2$residuals,type='l')

> m3=arima(c3,xreg=c1,order=c(0,0,1)) % Residuals follow an MA(1) model
> m3
Call:
  arima(x = c3, order = c(0, 0, 1), xreg = c1)
Coefficients:             ma1 intercept c1  % Fitted model is
  0.2115 0.0002  0.7824  % c3 = 0.0002 + 0.782c1 + a(t) + 0.212a(t-1)
s.e.  0.0224  0.0018  0.0077  % with var[a(t)] = 0.00446.
sigma^2 estimated as 0.004456:  log likelihood = 2531.84,  aic = -5055.69

> acf(m3$residuals)
> tsdiag(m3)

> m4=arima(c3,xreg=c1,order=c(1,0,0)) % Residuals follow an AR(1) model.
> m4
Call:
arima(x = c3, order = c(1, 0, 0), xreg = c1)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>ar1</th>
<th>intercept</th>
<th>c1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar1</td>
<td>0.1922</td>
<td>0.0003</td>
<td>0.7829</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.0221</td>
<td>0.0019</td>
<td>0.0077</td>
</tr>
</tbody>
</table>

% Fitted model is

\[ c3 = 0.0003 + 0.783c1 + a(t), \]

\[ a(t) = 0.192a(t-1)+e(t). \]

\[ \sigma^2 \text{ estimated as } 0.004474: \text{ log likelihood } = 2527.86, \ \text{aic } = -5047.72 \]

**Remark: Parameterization in R.** With additional explanatory variable \(X\) in ARIMA model, \(R\) use the model

\[ W_t = \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \]

where \(W_t = Y_t - \beta_0 - \beta_1 X_t\). This is the proper way to handle regression model with time series errors, because \(W_{t-1}\) is not subject to the effect of \(X_{t-1}\).

It is different from the model

\[ Y_t = \beta_0^* + \beta_1^* X_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \]

for which the \(Y_{t-1}\) contains the effect of \(X_{t-1}\).

**Long-memory models**

- Meaning? ACF decays to zero very slowly!

- Example: ACF of squared or absolute log returns

  ACFs are small, but decay very slowly.

- How to model long memory? Use “fractional” difference: namely,

  \[(1 - B)^d r_t, \text{ where } -0.5 < d < 0.5.\]

- Importance? In theory, Yes. In practice, yet to be determined.
• In R, the package fArma may be used to estimate the fractionally integrated ARMA models, but it requires certain Ox functions to be installed in some specific directories.

Summary of the chapter

• Sample ACF ⇒ MA order

• Sample PACF ⇒ AR order

• Some packages have “automatic” procedure to select a simple model for “conditional mean” of a FTS, e.g., R uses “ar” for AR models.

• Check a fitted model before forecasting, e.g. residual ACF and hetroscedasticity (chapter 3)

• Interpretation of a model, e.g. constant term &
  For an AR(1) with coefficient $\phi_1$, the speed of mean reverting as measured by half-life is
  \[ k = \frac{\ln(0.5)}{\ln(|\phi_1|)}. \]
  For an MA($q$) model, forecasts revert to the mean in $q+1$ steps.

• Make proper use of regression models with time series errors, e.g. regression with AR(1) residuals
  Perform a joint estimation instead of using any two-step procedure, e.g. Cochrane-Orcutt (1949).
Example: Is there a Friday effect on asset returns?

If a daily market index is used, serial correlation may exist.

- Basic properties of a random-walk model
- Multiplicative seasonal models, especially the so-called airline model.