Modeling and Forecasting Seasonality

Dr. Natalia Khorunzhina

CBS

Objective and Format of the Class

- Construct models with deterministic seasonality;
- Forecasting based on deterministic seasonality;
- Alternative way of modeling seasonality through trigonometric series;
- Application: modeling and forecasting hotel occupancy;
- Homework exercise: modeling and forecasting taxable revenue for real estate.

Forecast Period

- In-sample observations: $y_1, y_2, ..., y_T$
- Out-of-sample period: $y_{T+1}, y_{T+2}, ..., y_{T+h}$
- h is called the forecast horizon

Actual Forecasting

- Even if the variables in the information set Ω_t are known, the conditional mean function $E(y_{t+h}|\Omega_t)$ is unknown.
  - The functional form is unknown.
  - The parameters of the function are unknown.
- Thus to make an actual forecast, we need to:
  - Create an approximate model for $E(y_{t+h}|\Omega_t)$.
  - Estimate the model parameters from data.

Time-Series Components

- Recall that the optimal point forecast of a series y_{t+h} is its conditional mean:

$$\mu_t = E(y_{t+h}|\Omega_t)$$

- It is useful to decompose this mean into components:

$$\mu_t = T_t + S_t + C_t$$

  - T_t: Trend
  - S_t: Seasonal
  - C_t: Cycle

Components

- Trend
  - Very long term (decades)
  - Smooth
- Seasonal
  - Patterns which repeat annually
  - May be constant or variable
- Cycle
  - Business cycle
  - Correlation over 2-7 years
- It is useful to consider the components separately

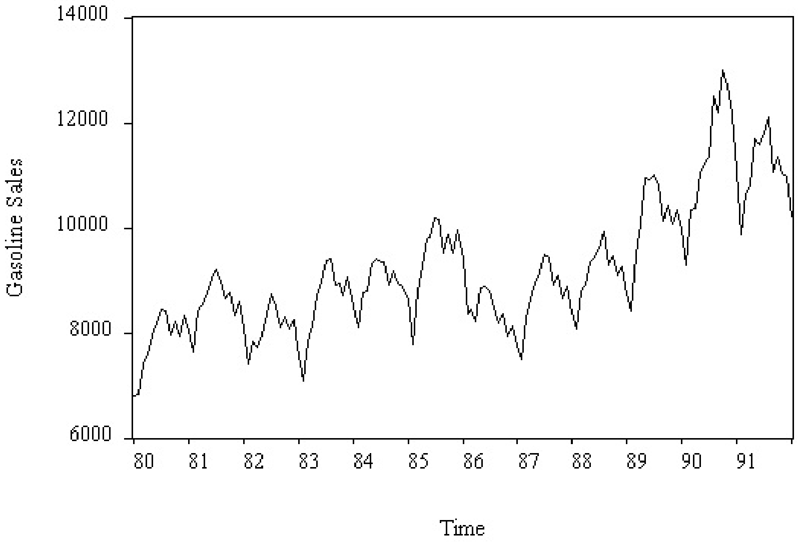
Seasonality

$$\mu_t = T_t + S_t + C_t$$

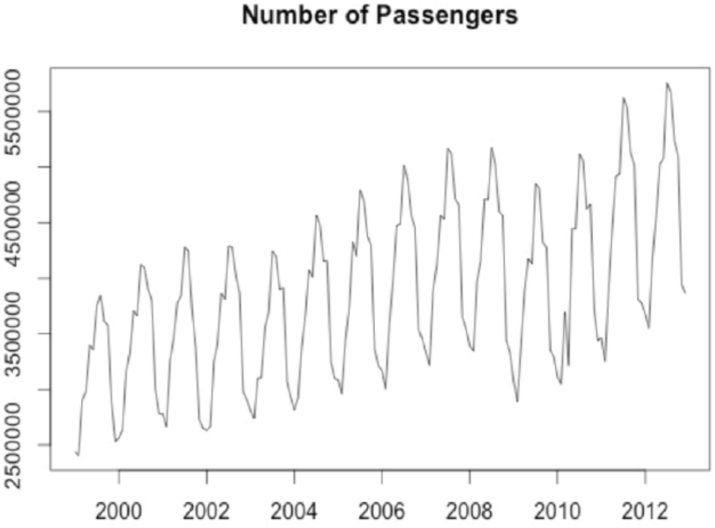
where S_t is the seasonal component.

- The seasonal component S_t is a repetitive cycle over the calendar year
- Seasonality S_t can be deterministic (predictable) or stochastic

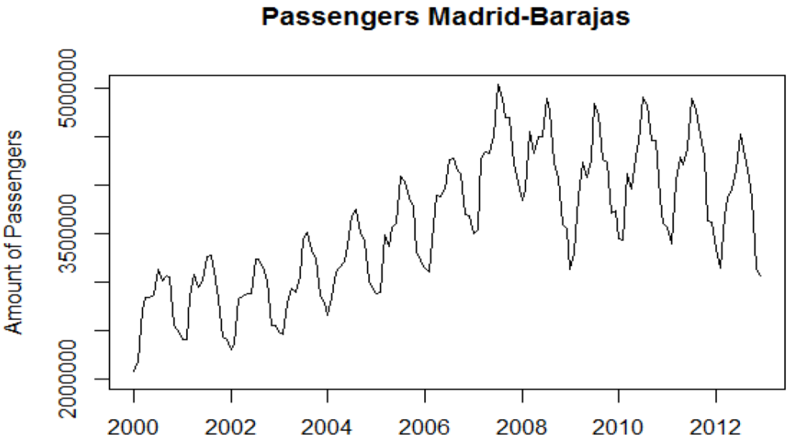
Gasoline Sales



Number of Passengers in Dutch Airports

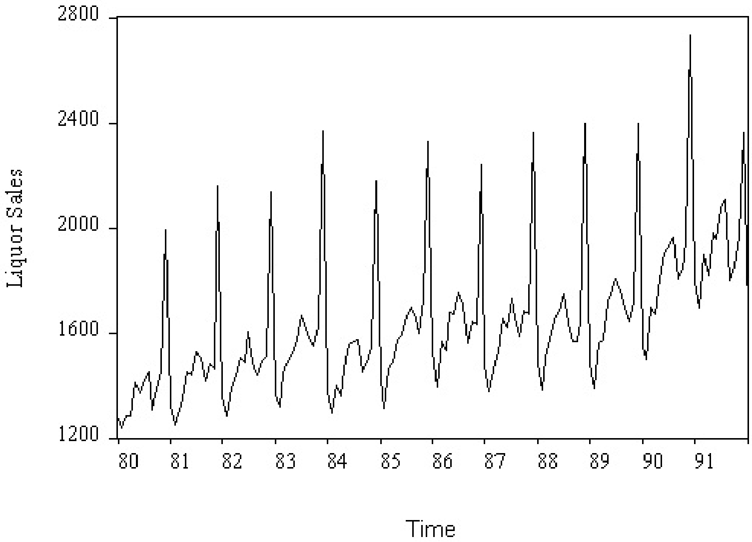


Number of Passengers in Madrid Airport

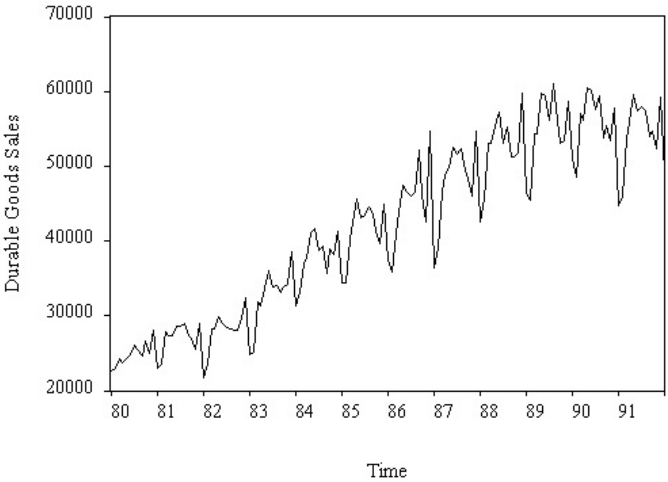


Data passengers Madrid-Barajas 2000-2012

Liquor Sales



Durable Goods Sales



Seasonality Examples

- Gasoline consumption rises in summer due to increased auto travel
- International airline prices rise in summer due to increased tourism
- Natural gas consumption and prices rise in winter due to heating
- Electricity consumption increases in summer due to air conditioning
- Construction activity and jobs decrease in winter in the Midwest
- Consumer spending increases in November and December due to holiday shopping

Deterministic vs Stochastic Seasonality

- If the seasonal pattern repeats year after year, it is deterministic and predictable.
  - Christmas is always in December
- If the seasonal pattern roughly repeats itself, but evolves over the years, it is stochastic and only partially predictable.
  - Holiday shopping as a percentage of income is not a fixed constant
- Seasonal patterns can change dramatically as the economy evolves.
  - The spread of air conditioning shifted the seasonal pattern of residential electricity consumption from winter to summer

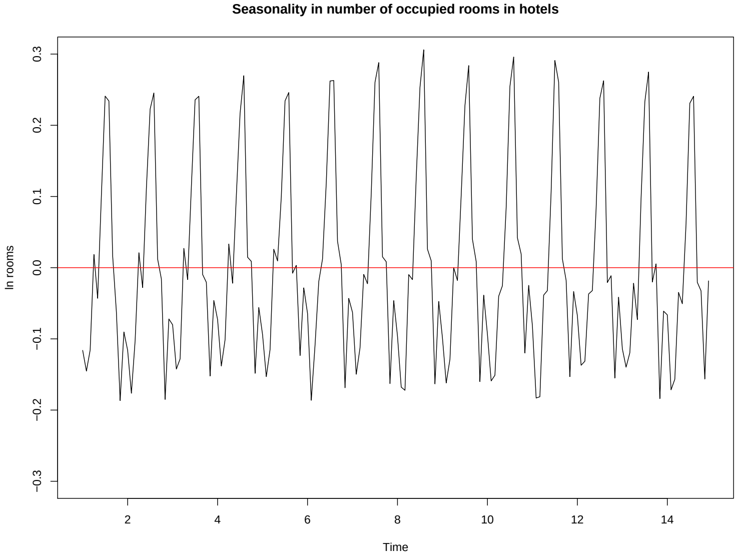
Deterministic Seasonality

- If seasonality is constant and deterministic the S_t is simply a different constant for each period
- For example, for monthly data:

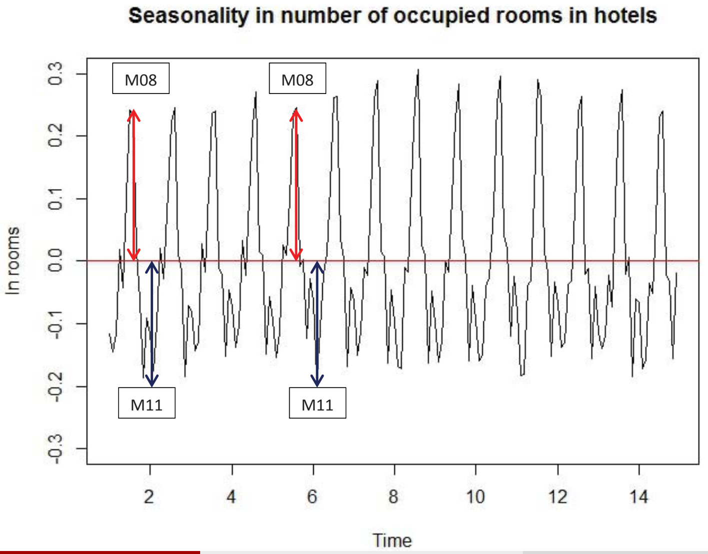
$$S_t = \begin{cases} \gamma_1 & \text{if } t = \text{January} \\ \gamma_2 & \text{if } t = \text{February} \\ \vdots & \qquad\vdots \\ \gamma_{12} & \text{if } t = \text{December} \end{cases}$$

- Seasonality is a constant which varies by the calendar period (quarter, month, week, day, or time of day)

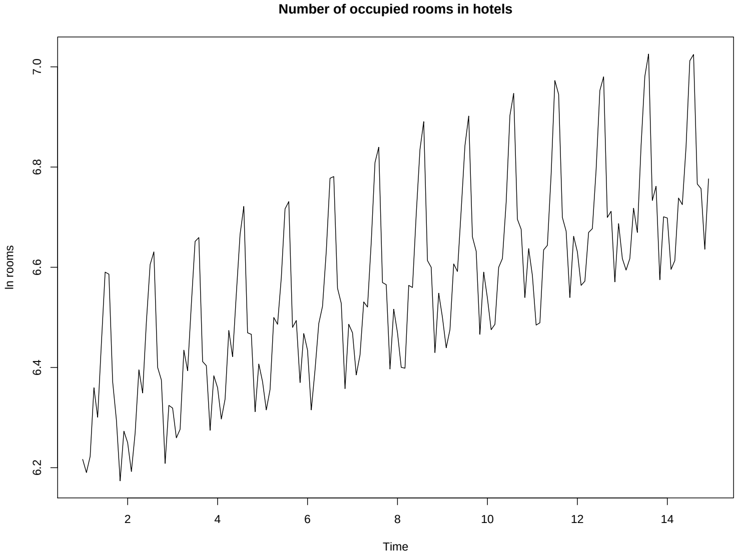
Here is the idea: in the detrended data



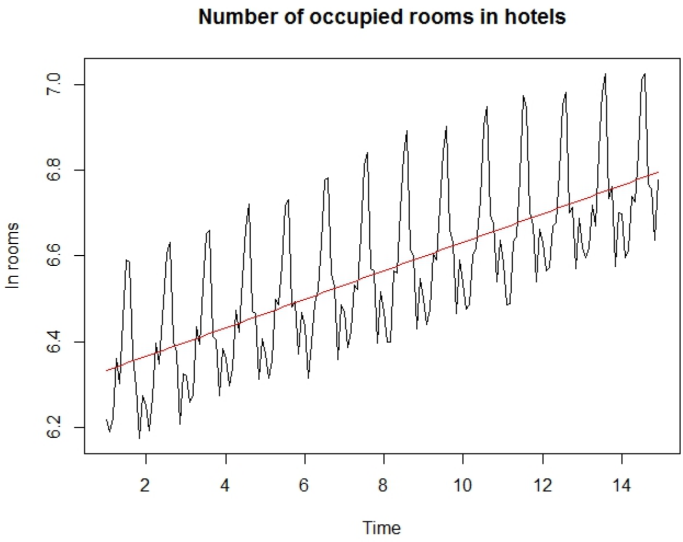
Here is the idea: in the detrended data



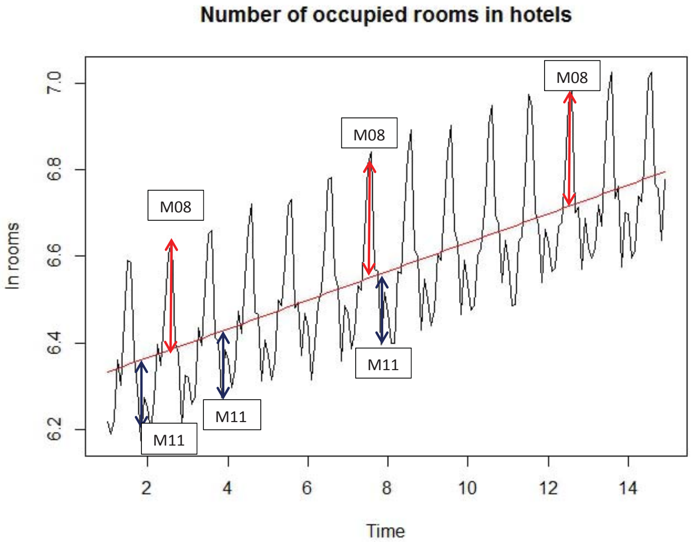
Here is the idea using hotel occupancy data



Here is the idea: with the trend



Here is the idea: with the trend



Seasonal Dummy Model

- Deterministic seasonality S_t can be written as a function of seasonal dummy variables
- Let s be the seasonal frequency
  - $s = 4$ for quarterly
  - $s = 12$ for monthly
- Let $D_1, D_{2t}, D_{3t}, ..., D_{st}$ be seasonal dummies
  - $D_{1t} = 1$ if s is the first period, otherwise $D_{1t} = 0$
  - $D_{2t} = 1$ if s is the second period, otherwise $D_{2t} = 0$
- At any time period t, one of the seasonal dummies $D_{1t}, D_{2t}, D_{3t}, ..., D_{st}$ will equal 1, all the others will equal 0.

Seasonal Dummy Variables with Quarterly Data

$$
\begin{aligned}
D_1 &= (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, \ldots) \\
D_2 &= (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, \ldots) \\
D_3 &= (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \ldots) \\
D_4 &= (0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, \ldots)
\end{aligned}
$$

Seasonal Dummy Model

- Deterministic seasonality:

$$
\begin{aligned}
S_t &= \begin{cases} \gamma_1 & \text{if } t = \text{January} \\ \gamma_2 & \text{if } t = \text{February} \\ \vdots & \vdots \\ \gamma_{12} & \text{if } t = \text{December} \end{cases} \\
&= \sum_{i=1}^{12} \gamma_i D_{it}
\end{aligned}
$$

A linear function of the dummy variables.

Estimation

- Least squares regression:

$$
\begin{aligned}
y_t &= \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t \\
&= \alpha + \sum_{i=1}^{s-1} \beta_i D_{it} + e_t
\end{aligned}
$$

- You can either:
  - Regress y on all the seasonal dummies, omitting the intercept, or
  - Regress y on an intercept and the seasonal dummies, omitting one dummy (one season, e.g. December)
- You cannot regress on both the intercept plus all seasonal dummies, for they would be collinear and redundant.

Interpreting Coefficients

- In the model

$$S_t = \alpha + \sum_{i=1}^{s-1} \beta_i D_{it}$$

  the intercept $\alpha = \gamma_s$ is the seasonality in the omitted season.
- The coefficients $\beta_i = \gamma_i - \gamma_s$ are the difference in the seasonal component from the s'th period.

Model with Trend and Seasonality

Trend may be included as well, in which case the model is:

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

Seasonal Adjustment

- Most economic indicators reported by the government are **seasonally adjusted.**
- Roughly, the component S_t is estimated, and then what is reported is

$$\begin{aligned} y_t^* &= y_t - S_t \\ &= T_t + C_t \end{aligned}$$

- The idea is that seasonality distracts from the main reporting purpose
  - Seasonally adjusted data allows users to focus on trend and business cycle movements
- Seasonal adjustment by central statistical agencies is sophisticated, allowing for evolving seasonal patterns.

Forecasting Trend and Seasonal Components

- Consider a point forecast. Suppose we are at time T and we want to forecast the h-step-ahead value of a series y_t.
- The ”true” model is:

$$y_t = \beta_1 \; TIME_t + \sum_{i-1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- Then, at time $T + h$,

$$y_{T+h} = \beta_1 \; TIME_{T+h} + \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \varepsilon_{T+h}$$

- Optimal forecast of ε_{T+h}, assuming $\varepsilon_t \sim IID(0, \sigma^2)$:

$$\hat{y}_{T+h,T} = \hat{\beta}_1 \; TIME_{T+h} + \sum_{i=1}^{s} \hat{\gamma}_i D_{i,T+h}$$

Forecasting Trend and Seasonal Components

- Consider now an interval forecast. We assume that the regression disturbance is normally distributed.
- A 95% interval forecast, ignoring parameter estimation uncertainty, is:

  $$y_{T+h} \pm 1.96\sigma$$

  where σ is the standard deviation of the disturbance in the regression.
- We use the forecast:

  $$\widehat{y}_{T+h} \pm 1.96\hat{\sigma}$$

  where $\hat{\sigma}$ is an estimate of σ.

Trigonometric seasonality models

Sometimes regression models involving **trigonometric terms** can be used to forecast time series. Let s be the seasonal frequency of the data.

- $y_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{s}\right) + \beta_3 \cos\left(\frac{2\pi t}{s}\right) + \varepsilon_t$
- $y_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{s}\right) + \beta_3 \cos\left(\frac{2\pi t}{s}\right) + \beta_4 \sin\left(\frac{4\pi t}{s}\right) + \beta_5 \cos\left(\frac{4\pi t}{s}\right) + \varepsilon_t$

The first model is useful in modeling a very regular seasonal series with constant seasonal variation.

The second model possesses terms that allow for modeling of time series displaying constant seasonal variation and having more complicated seasonal pattern.