MS&E 125: Intro to Applied Statistics

Linear regression

Professor Udell

Management Science and Engineering
Stanford

May 31, 2023

source: Modified from Peter Frazier, Cornell ORIE 3120

Outline

Forecasting time series

A time series, $x_1, x_2, x_3, \ldots$ is a data sequence observed over time, for example,

- demand for parts
- sales of a product
- unemployment rate

we'll study special methods for forecasting time series.

- develop an algorithm to track the time series and to extrapolate into the future

Outline

Constant mean model: introduction

Suppose demand for a product follows the (very) simple model

$$x_n = \mu + \varepsilon_n$$

Here

- x_n = demand for time period n
- μ is the expected demand – constant in this simple model
- $\varepsilon_1, \varepsilon_2, \ldots$ are independent with mean 0
- the best forecast of a future value of x_n is μ
- we want to estimate μ and update the estimate as each new x_n is observed

Constant mean model: forecasts

Suppose we estimate the demand as an average of the observed values

$$\hat{\mu}_n = \frac{x_1 + \cdots + x_n}{n}$$

Constant mean model: forecasts

Suppose we estimate the demand as an average of the observed values

$$\hat{\mu}_n = \frac{x_1 + \cdots + x_n}{n}$$

Define $\hat{x}_n(\ell)$ to be the ℓ-step ahead forecast at time period n

- $\hat{x}_n(\ell)$ is the forecast at time n of demand at time $n + \ell$
- forecast might change as we observe more data

Constant mean model: forecasts

Suppose we estimate the demand as an average of the observed values

$$\hat{\mu}_n = \frac{x_1 + \cdots + x_n}{n}$$

Define $\hat{x}_n(\ell)$ to be the ℓ-step ahead forecast at time period n

- $\hat{x}_n(\ell)$ is the forecast at time n of demand at time $n + \ell$
- forecast might change as we observe more data

Then, in this simple model, the best forecasts at time n are

$$\hat{x}_n(\ell) = \hat{\mu}_n, \text{ for all } \ell > 0$$

Constant mean model: updating $\hat{\mu}_n$

In this simple model, μ does not change, but our estimate of μ does

Constant mean model: updating $\hat{\mu}_n$

In this simple model, μ does not change, but our estimate of μ does

easy to update $\hat{\mu}_n$ to $\hat{\mu}_{n+1}$:

$$\begin{aligned}
\hat{\mu}_{n+1} &= \frac{(x_1 + \cdots + x_n) + x_{n+1}}{n+1} \\
&= \frac{n}{n+1}\hat{\mu}_n + \frac{1}{n+1}x_{n+1} \\
&= \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)
\end{aligned}$$

Advantages of the updating formula

The simple updating formula

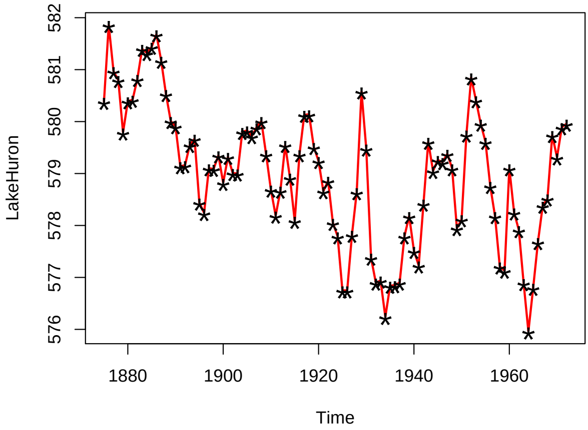
$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$$

has several advantages:

- reduced storage
  - we only store $\hat{\mu}_n$
- computational speed
  - the mean need not be recomputed each time
- suggests ways to handle a slowly changing mean
  - coming soon

Outline

Lake Huron level – example with a slowly changing mean



Slowly changing mean model: introduction

- Now suppose that

$$x_n = \mu_n + \varepsilon_n$$

  where μ_n is slowly changing

- The forecast is the same as for the constant mean model:

$$\hat{x}_n(\ell) = \hat{\mu}_n, \text{ for all } \ell > 0$$

- What changes is the way $\hat{\mu}_n$ is updated
  - We need $\hat{\mu}_n$ to track μ_n

Slowly changing mean: updating

- For a constant mean, the update is

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$$

- For a slowly changing mean, the update is

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \alpha(x_{n+1} - \hat{\mu}_n) = (1 - \alpha)\hat{\mu}_n + \alpha x_{n+1}$$

  for a constant α

- α is adjusted depending on how fast μ_n is changing
  - $0 < \alpha < 1$
  - faster changes in μ necessitate larger α

Demo: Exponential smoothing

`https://colab.research.google.com/github/stanford-mse-125/demos/blob/main/forecasting.ipynb`

Exponential weighting

Start with the updating equation and iterate backwards:

$$
\begin{aligned}
\hat{\mu}_{n+1} &= (1-\alpha)\hat{\mu}_n + \alpha x_{n+1} \\
&= (1-\alpha)\{\hat{\mu}_{n-1}(1-\alpha) + \alpha x_n\} + \alpha x_{n+1} \\
&= (1-\alpha)^2\hat{\mu}_{n-1} + (1-\alpha)\alpha x_n + \alpha x_{n+1} \\
&= (1-\alpha)^3\hat{\mu}_{n-2} + (1-\alpha)^2\alpha x_{n-1} + (1-\alpha)\alpha x_n + \alpha x_{n+1} \\
&\approx \alpha\Big\{x_{n+1} + (1-\alpha)x_n + (1-\alpha)^2 x_{n-1} \\
&\qquad +(1-\alpha)^3 x_{n-2} + \cdots + (1-\alpha)^n x_1\Big\}
\end{aligned}
$$

Hence $\hat{\mu}_{n+1}$ is an exponentially weighted moving average Large values of α mean faster discounting of the past values.

Exponential weighted moving average (math)

Use previous page + summation formula for geometric series (see next page):

$$\hat{\mu}_{n+1} \approx$$

$$\alpha\left\{(1-\alpha)^0 x_{n+1} + (1-\alpha)^1 x_n + (1-\alpha)^2 x_{n-1} + \cdots + (1-\alpha)^n x_1\right\}$$

$$\approx \frac{\left\{(1-\alpha)^0 x_{n+1} + (1-\alpha)^1 x_n + (1-\alpha)^2 x_{n-1} + \cdots + (1-\alpha)^n x_1\right\}}{1 + (1-\alpha) + \cdots + (1-\alpha)^n}$$

Summing a geometric series (math)

Assume $|\gamma| < 1$ so $\gamma^n \to 0$ as $n \to \infty$

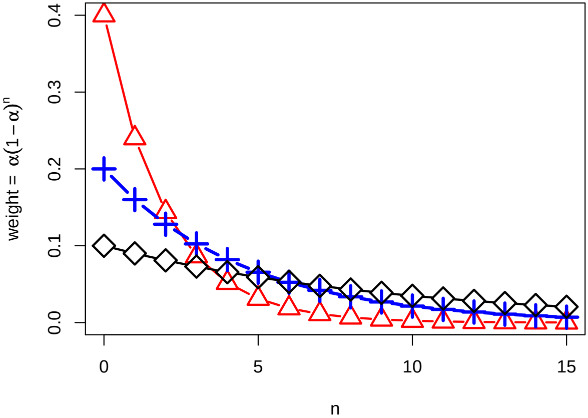
$$1 + \gamma + \gamma^2 + \cdots + \gamma^n = \frac{1 - \gamma^{n+1}}{1 - \gamma} \approx \frac{1}{1 - \gamma} \text{ (if } n \text{ is large enough)}$$

Now let $\gamma = 1 - \alpha$. Then

$$1 + (1 - \alpha) + \cdots + (1 - \alpha)^n \approx \frac{1}{\alpha}$$

since $1 - (1 - \alpha) = \alpha$.

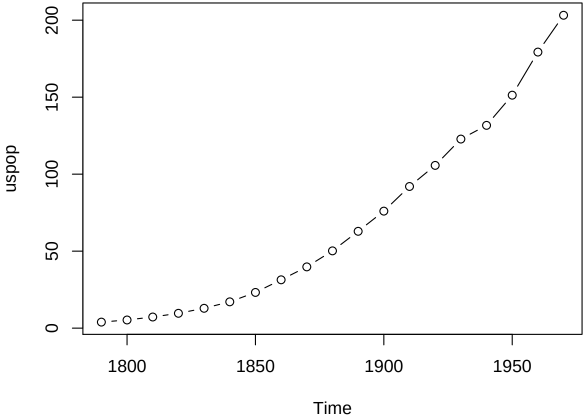
Exponential weights: examples



Note: weights start at α when $n = 0$

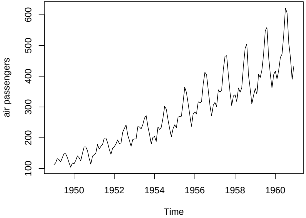
Outline

Forecasting with trends: example



Census counts of US population

Forecasting with trends and seasonality: example



Airline passenger miles

Note the seasonal pattern and trend in this example

- typical of business data

Holt method: forecasting with trend

For now, assume data has trend but no seasonality

Holt's forecasting method uses a linear trend

$$\text{estimate at time } n \text{ of } x_{n+\ell} := \hat{x}_n(\ell) = \hat{\mu}_n + \hat{\beta}_n \ell$$

- n is "origin" – time when forecasts are being made
- ℓ is the "lead" – how far ahead one is forecasting
- $\hat{\mu}_n$ is called the level
- $\hat{\beta}_n$ is called the slope

Both $\hat{\mu}_n$ and $\hat{\beta}_n$ are updated as we make more observations n

Holt method: Updating the level

In the Holt model, the level $\hat{\mu}_n$ is updated by the equation:

$$\hat{\mu}_{n+1} = (1-\alpha)(\hat{\mu}_n + \hat{\beta}_n) + \alpha x_{n+1}$$

or, equivalently,

$$\hat{\mu}_{n+1} = \hat{\mu}_n + (1-\alpha)\hat{\beta}_n + \alpha(x_{n+1} - \hat{\mu}_n)$$

- $\hat{\mu}_n + \hat{\beta}_n$ is predicted value at time $n+1$
- α is for updating the level and β for the slope (next)

Compare with previous update equation (for no-trend model):

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \alpha(x_{n+1} - \hat{\mu}_n) = (1-\alpha)\hat{\mu}_n + \alpha x_{n+1}$$

Holt model: updating the slope

In the Holt model, the slope $\hat{\beta}_n$ is updated by the equation:

$$\hat{\beta}_{n+1} = (1 - \beta)\hat{\beta}_n + \beta(\hat{\mu}_{n+1} - \hat{\mu}_n)$$

or, equivalently,

$$\hat{\beta}_{n+1} = \hat{\beta}_n + \beta \left\{ (\hat{\mu}_{n+1} - \hat{\mu}_n) - \hat{\beta}_n \right\}$$

Demo: Holt's method

https://colab.research.google.com/github/stanford-mse-125/demos/blob/main/forecasting.ipynb

Winters' additive seasonal method

Winters extended Holt's method to include seasonality. The method is usually called Holt-Winters forecasting

Let s be the period length:

- $s = 4$ for quarterly data
- $s = 12$ for monthly data
- $s = 52$ for weekly data
- $s = 13$ for data collected over 4-week periods
- $s = 24$ for hourly data

Holt-Winters updating

Holt-Winters forecasting can use either of two types of updating

- additive
- multiplicative

These refer to how the trend and seasonal components are put together

- the trend and seasonal components can be added or multiplied

Holt-Winters additive seasonal method

The forecasts are periodic

With the additive methods they are:

$$
\begin{aligned}
\hat{x}_n(\ell) &= \hat{\mu}_n + \hat{\beta}_n \ell + \widehat{S}_{n+\ell-s}, \text{ for } \ell = 1, 2, \ldots, s \\
&= \hat{\mu}_n + \hat{\beta}_n \ell + \widehat{S}_{n+\ell-2s}, \text{ for } \ell = s+1, \ldots, 2s
\end{aligned}
$$

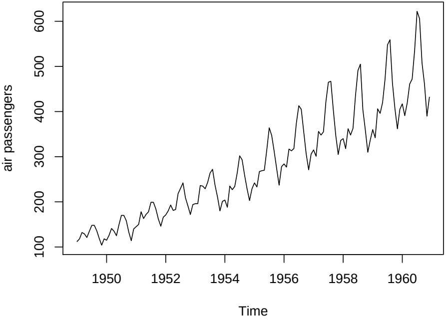
and so forth

Winters' additive seasonal model: updating

$$\begin{aligned}\hat{\mu}_{n+1} &= \alpha(x_{n+1} - \widehat{S}_{n+1-s}) + (1-\alpha)(\hat{\mu}_n + \hat{\beta}_n) \\ \hat{\beta}_{n+1} &= \beta(\hat{\mu}_{n+1} - \hat{\mu}_n) + (1-\beta)\hat{\beta}_n \\ \widehat{S}_{n+1} &= \gamma(x_{n+1} - \hat{\mu}_{n+1}) + (1-\gamma)\widehat{S}_{n+1-s}\end{aligned}$$

α, β, and γ are "tuning parameters" that we need to adjust

Why we need multiplicative seasonal models



Notice the multiplicative behavior

- the seasonal fluctuations are larger where the trend is larger

Holt-Winters multiplicative seasonal method

$$
\begin{aligned}
\hat{x}_n(\ell) &= (\mu_n + \hat{\beta}_n \ell)\widehat{S}_{n+\ell-s}, \text{ for } \ell = 1, 2, \ldots, s \\
&= (\mu_n + \hat{\beta}_n \ell)\widehat{S}_{n+\ell-2s}, \text{ for } \ell = s+1, \ldots, 2s
\end{aligned}
$$

and so forth

Winters' multiplicative seasonal model: updating

$$
\begin{aligned}
\hat{\mu}_{n+1} &= \alpha \frac{x_{n+1}}{\widehat{S}_{n+1-s}} + (1-\alpha)(\hat{\mu}_n + \hat{\beta}_n) \\
\hat{\beta}_{n+1} &= \beta(\hat{\mu}_{n+1} - \hat{\mu}_n) + (1-\beta)\hat{\beta}_n \\
\widehat{S}_{n+1} &= \gamma \frac{x_{n+1}}{\hat{\mu}_{n+1}} + (1-\gamma)\widehat{S}_{n+1-s}
\end{aligned}
$$

Demo: Exponential smoothing

`https://colab.research.google.com/github/stanford-mse-125/demos/blob/main/forecasting.ipynb`

Applications:

- Lake Huron
- US population
- CO2
- Airline passengers
- Sales

Outline

Residuals

For given values of α, β, and γ:

- $\hat{\mu}_n, \hat{\beta}_n, \widehat{S}_n, \ldots, \widehat{S}_{n-s}$ are the level, slope, and seasonalities at time n
- $\hat{x}_{n+1} = x_n(1) = \hat{\mu}_n + \hat{\beta}_n + \widehat{S}_{n+1-s}$ is the one-step ahead forecast at time n
- $\hat{\epsilon}_{n+1} = x_{n+1} - \hat{x}_{n+1}$ is the residual or one-step ahead forecast error

Choosing α, β, and γ

α, β, and γ are called "tuning parameters"

Suppose we have data $x_1, \ldots, x_N$:

- the usual way to select α, β, and γ is to minimize

$$\text{SS}(\alpha, \beta, \gamma) = \sum_{n=N_1+1}^{N} \hat{\epsilon}_n^2$$

  where the first N_1 residuals are discarded to let the forecasting method "burn-in"
- this technique is used by statsmodels, unless the user specifies the parameters explicitly in the `fit` call

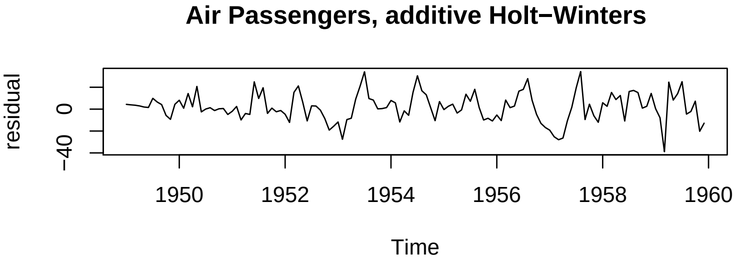
Comparing forecasting methods and diagnosing problem

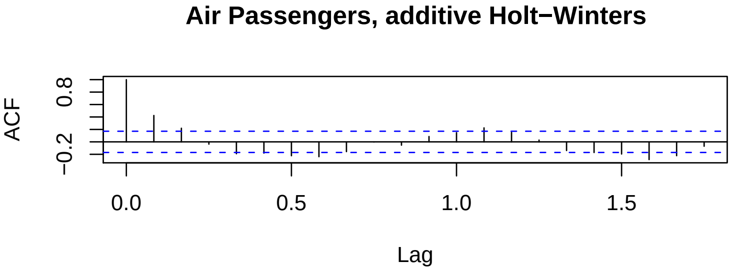
- Two or more forecasting methods can be compared using

$$\min_{\alpha,\beta,\gamma} \text{SS}(\alpha, \beta, \gamma)$$

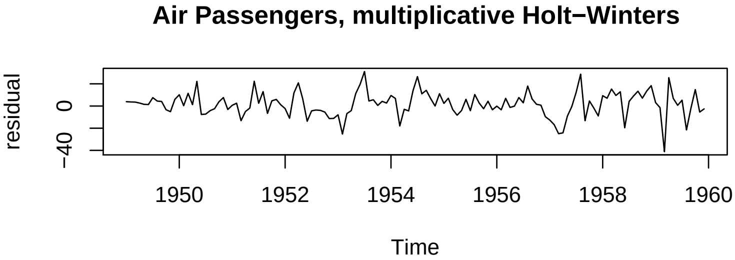
- If a forecasting method is working well, then the residuals should not exhibit autocorrelation

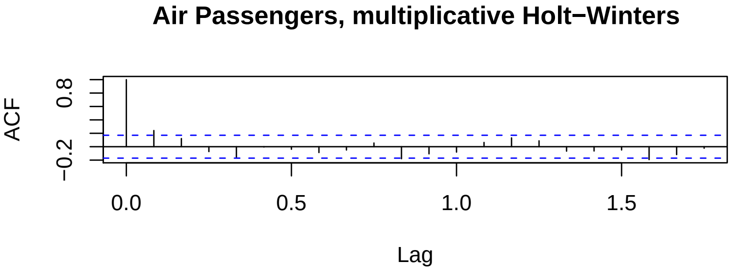
Air passengers: additive seasonal method





Air passengers: multiplicative seasonal method





Outline

Residuals

Selecting the tuning parameters

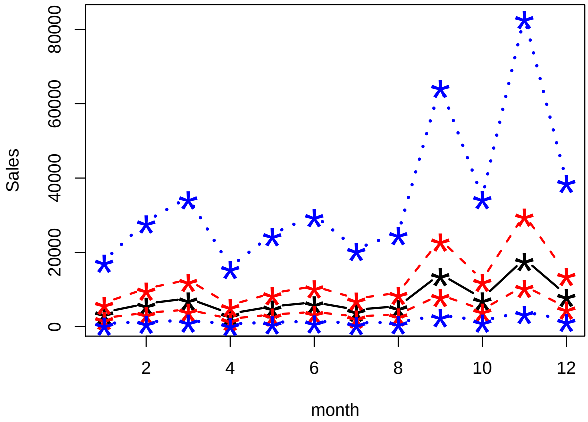
Forecasting using regression

Forecasting using regression

In some situations, regression can be used for forecasting

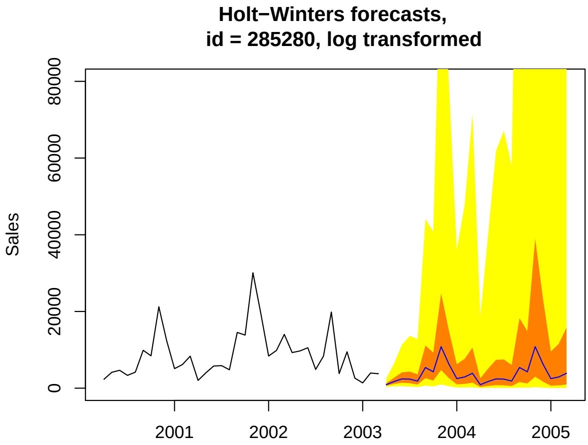
- in the following example, regression will be used to forecast Stove Top product 285280
  - this is the product that we forecast earlier with Holt-Winters
- the regression model will have seasonal effects but not trend
  - the seasonal effects will be introduced by using `month` as a factor
- regression uses all the data to estimate the level and the seasonal effects
  - so there is no discounting of the past
  - this helps us deal with the small amount of data

Stove Top product 285280 forecasts using regression



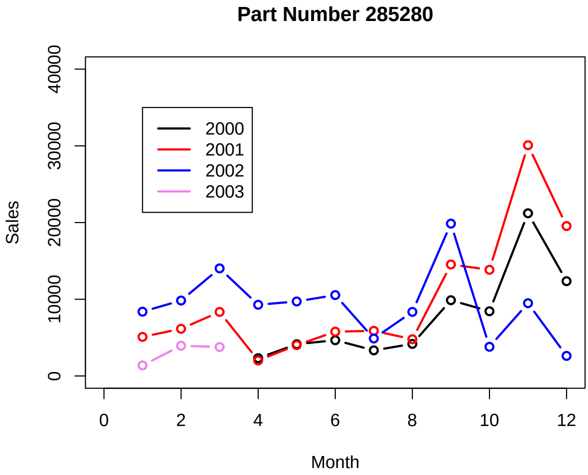
black = predictions, red = 50% pred. int., blue = 95% pred. int.

Holt-Winters product 285280 forecasts: log transformed, zoom in



For comparison, here again are the forecasts from Holt-Winters.

Why is forecasting so difficult with this product?



Sales patterns vary across years. e.g., in 2002, holiday sales came earlier.