Linear Regression w/in A Time Series Context

Shpaner, Leon

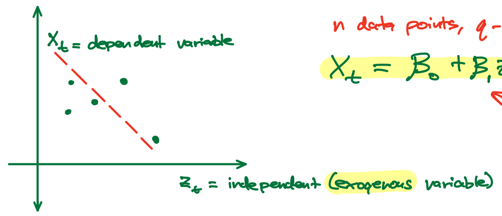
quantify - not just sum of squares error

- compare models
- reduced model
- Exploratory Data Analysis

Stationarity : 2 different things you can do w/ data that's not stationary

① trend
② differencing

* PAY CLOSE ATTENTION TO NOTATION USED IN BOOK



n data points, q - variables

$$X_t = \beta_0 + \beta_1 z_1 t + \beta_2 z_2 t$$

independent

- we want to predict x from z, that's why x is on the vertical axis

How do we fit this model?

hat means estimator

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^{n} (z_t - \bar{z})^2}$$

$$\beta_0 = \underbrace{\bar{x}}_{\text{mean of } x} - \hat{\beta}_1 \bar{z}$$

How well it fits:

$$\sum_{t=1}^{n} (\underbrace{x_t}_{\text{actual points}} - \underbrace{\hat{x}_t}_{\text{prediction (dotted line)}})^2 = SSE$$

$$\frac{SSE}{n(q+1)}$$

q: # of independent variables

Sum squared errors / # of deg. of freedom → mean-squared error

t - distribution which lends itself to hypothesis testing.

Reduced Model

q = # of variables

r = # of variables in reduced model

- now we can take the sum of squared errors from the original (full) model, and the reduced model, and just compare SSE directly.
  ↳ these will give us a distribution that's governed by an f-Statistic which is telling us which specific statistical tests we can use to say whether or not there is a significant difference in how well two models fit, and how well these two models can explain the data.

F - Statistic

$$F = \frac{MSR}{MSE} = \frac{(SSE_r - SSE)/(q-r)}{SSE/(n-q-1)}$$

in statistics, this would come in the form of an ANOVA table telling you how to interpret these quantities.

A model w/ no structure, with just a mean & white noise would have an

$r = 0$

$$x_t = \beta_0 + \underbrace{w_t}_{\text{white noise}}$$

there's no time structure to this model, it's just white noise. You would hope that adding in any of your independent var. should fit better than this. Definitely something you want to be able to calculate that sum of squared errors for, or a null model → you should always make sure that your model fits better than assuming that the data are just noise.

If you write the sum of squares model where it's just noise or the sum of squares model where R is zero, it's just: $SSE_0 \sum_{t=1}^{n} (x_t - \bar{x}) = SST$

$$\frac{SSR}{SST} = R^2$$

R^2 is just another way of saying how good a model is. It's going to be closely related to those things like correlation, covariance, and slope. R^2 is just a quantity that has almost all the same terms in it.

- As you make your model more complicated, it's going to fit better b/c it has to. If you're adding in new variable you can incorporate it (if it helps) or choose to ignore it as white noise.

As you add more variables to the model, you want to put a penalty so it pays for that complexity.

$\hat{\sigma}_K = \frac{SSE(K)}{n}$ ← # of data points

When you've obtained a couple of these quantities, you can now calculate from this a penalized version of how well the model fits.

What's the cost of adding more predictors to your model?

$$AIC = \underbrace{\log(\sigma_K)}_{\text{error}} + \underbrace{\frac{n+2K}{n}}_{\text{penalty}}$$

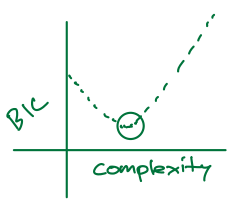
→ as your number of data points increases, this is going to increase. This term is basically your error. You want your error to be small. This is your penalty for building a more complicated model.

$$AIC_c = \underbrace{\log(\sigma_K)}_{\text{error}} + \underbrace{\frac{n+K}{n-K-2}}_{\text{penalty}}$$

→ Corrected AIC

$$BIC = \underbrace{\log(\sigma_K)}_{\text{error}} + \underbrace{\frac{K\log n}{n}}_{\text{penalty}}$$

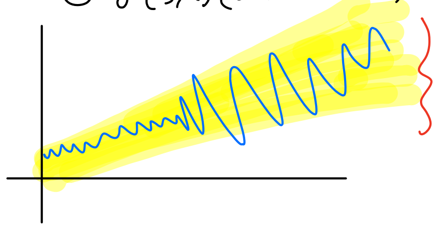
a couple of different ways of accomplishing the same task, which is taking your error and making it so that a more complicated model has to pay more, so you want your error to be small, but penalty is going up as model gets more complicated.



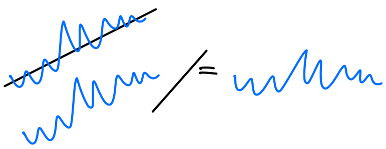
Linear Regression - Time Series

Stationarity -

1. μ is constant (doesn't depend on time).
2. $\gamma(s,t)$ (autocovariance) only depends on time difference. If your variance is increasing over time, this won't work.



clearly NOT Stationary b/c of change in variance.



trend stationarity : if there's some trend in the data & you remove the trend, you will then get something that's stationary.

fit the line, subtract the line → de-trending

Differencing : take our starting time series and transform it into a stationary time series, then fit the stationary time series.

$\nabla x_t = x_t - x_{t-1}$ ← differencing operation

$x_t = \underset{\substack{\uparrow\\ \text{trend}\\ \text{(not stationary)}}}{\mu_t} + \underset{\substack{\uparrow\\ \text{stationary}}}{Y_t}$ → When things are written w/out a hat, these are our theoretical quantities (what our model really is).

$\mu_t = \delta + \mu_{t-1} + w_t$ → plug in the random walk w/ drift

$x_t = (\mu_t + Y_t) - (\mu_{t-1} + Y_{t-1})$

$= \underset{\substack{\uparrow\\ \text{constant}}}{\delta} + \underset{\substack{\uparrow\\ \text{white noise}}}{w_t} + \underbrace{Y_t - Y_{t-1}}$

pair-wise difference btwn our input of variables. This in itself is stationary.
You can do the same thing w/ x above.

Define Backshift Operator:

$$B^k x_t = x_{t-k}$$

ex. $B^3 x_t = x_{t-3}$ → go back 3 samples

$B^{-1} x_{t-1} = B x_t$ → inverse

$B^{-1} B = 1$

$\nabla x = x_t - x_{t-1}$

$(1-B) x_t$

$(1-B^2) x = x_t - 2x_{t-1} + x_{t-2}$

$\nabla^2 = (1-B)^2$

Moving average

$$m_t = \sum_{j=-k}^{k} \alpha_j x_{t-j}$$

k: parameter of how big our average is

α_j: weights

x_{t-j}: lag back by some #

Kernel smoother → smoothing across the whole dataset

$$m_t = \sum_{i=1}^{n} w_i(t) x_{t_i}$$ → basically giving us smooth way to get coefficient