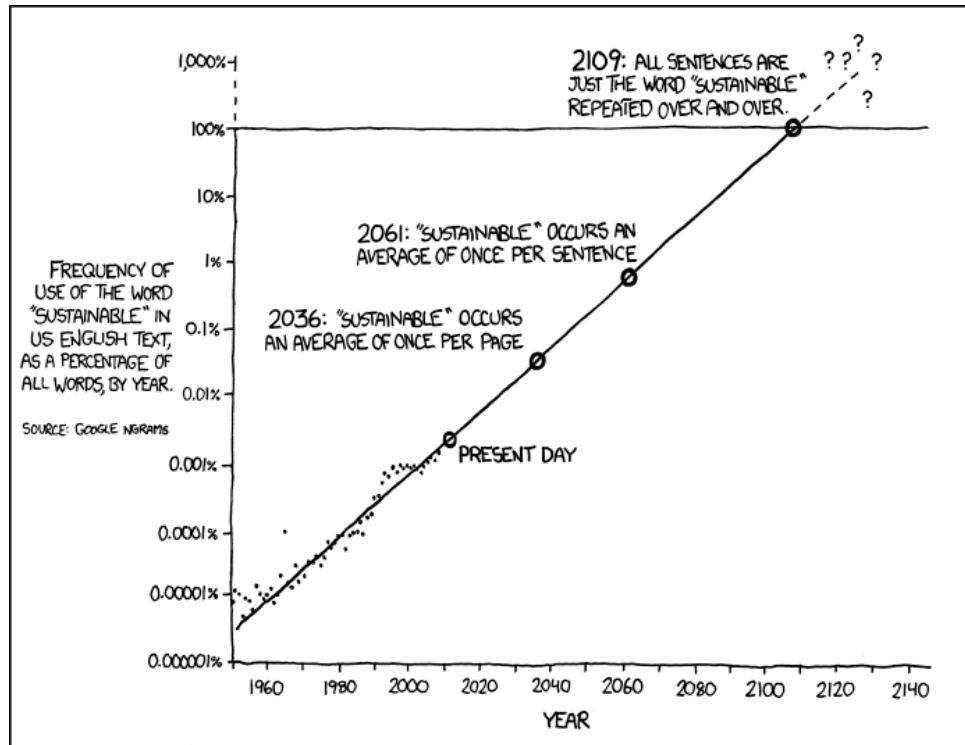


LECTURE 33: TIME SERIES AND FORECASTING

- I. Predicting the future
 - a. So far we've figured out how to estimate a relationship between several independent variables and one dependent variable.
 - b. We also discussed how to do a time series: where the data spans across time periods rather than places (e.g. countries and states).
 - c. One thing we can do is use this information to predict the future, either by:
 - i. Estimating future values of independent variables based on the year. For example, estimating the cost of Social Security by estimating how many elderly individuals there will be.
 - ii. Putting "year" into the regression directly as an independent variable. This is referred to as a *time series*.
 - iii. The former method is more accurate, but more complicated. The latter method is often sufficient for a general estimate of, say, sales or population.
 - iv. Note that the former method is just the latter method several times over.
 - d. To estimate the data, we just treat "year" like any other variable. If the coefficient for year is 0.2 and we want to know what the world will be like in 2025, we estimate by multiplying 0.2 and 2025. After considering the other variables (if any) and adding the intercept we have an estimate of the future.
 - e. This is called *trend projection*.
- II. Forecasting warnings
 - a. Forecasting is powerful, but not all-powerful. Like regressions, it comes with some assumptions, namely that the pattern of data won't change.
 - b. The problem to keep in mind with forecasting is naively forecasting. Consider this comic:



- i. Will every word we say be “sustainable” in 2109? Of course not, but the naïve forecast would suggest otherwise.
- c. This idea points to a more general concern: be wary of predicting values outside the range of your data. Predicting a few years ahead (or a few units beyond your maximum or minimum values) isn’t a big deal. But the farther you stray from what you know, the more likely you’ll look silly.

III. Autocorrelation

- a. Recall an assumption of regressions: residuals are independent from one another.
- b. When this assumption doesn’t hold—when residuals are not independent from one another—it’s called *autocorrelation*. It’s a particularly common problem in a time series.
 - i. Statistical significance is unreliable when there’s autocorrelation. It can cause Type I and Type II error.
- c. A test called the Durbin-Watson statistic can detect autocorrelation. We don’t have time to get into the details, but it basically measures how residuals changes from time period to time period to see if there’s a pattern.
- d. To way to solve autocorrelation is to add explanatory variables. By explaining the variation, you change the residual and, hopefully, eliminate the pattern.