MPRA
Munich Personal RePEc Archive

Fukushima: The Failure of Predictive Models

Stacey, Brian

Southern New Hampshire University

1 April 2015

Online at https://mpra.ub.uni-muenchen.de/69383/
MPRA Paper No. 69383, posted 10 Feb 2016 06:13 UTC

Fukushima

The Failure of Predictive Models

Brian Stacey

Southern New Hampshire University

Linear regression is an attempt to build a model from existing data. It uses the existing data to fit a linear equation that exhibits the least error from the actual points. Even within the existing data, a regression equation has limited ability to predict actual dependent variable values. Linear regression is limited in several ways; it assumes only linear relationships between variables, it is very sensitive to outliers, and data points must be independent. If any of these assumptions are not met the models get less accurate. Often random error (residuals) are enough to reduce the quality of fit of the equation to where it is unable to predict any values within the range of the original data except $(\bar{x}, \bar{y})$. Even with the best coefficient of determination (R^2=1), data within the sample set may not be indicative of the condition outside the sample set.

"The value of the dependent variable cannot be legitimately estimated if the value of the independent variable is outside the range of values in the sample data that served as the basis for determining the linear regression equation. There is no statistical basis to assume that the linear regression model applies outside of the range of the sample data.

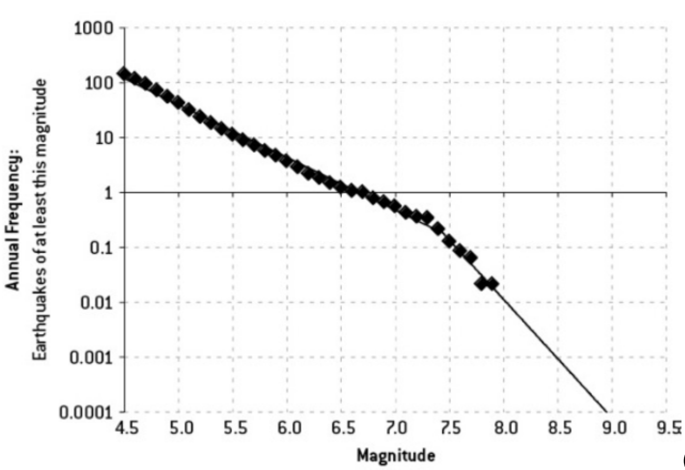
If the estimate of the dependent variable in fact concerns prediction, the historical data used to determine the regression equation might not be appropriate to represent future relationships. Unfortunately, one can only sample past data, not future data." (Kamer-Ainur, A., 2007).

Furthermore, even the strongest correlation does not mean causation. It might merely indicate that both the dependent and independent variables are equally affected by some third, unknown, factor. This would, at first glance appear to mean that estimation just outside of the sample data set would be accurate; if there is some outside factor that affects both variables equally, wouldn't it hold true that it would affect all values? This outside factor may not affect all values the same, and its influence may end at the end of your sample. A slight change in the slope of the actual affect near either end of the sample data would have an insignificant effect on the regression slope (β_1), however may be indicative of a change in the relationship between the variables beyond that point, conversely, a slight change in the slope of the sampled data near one end may be an artifact of sampling, or simply random error.

These weaknesses in linear regression are all magnified the further from the mean values ($\bar{x}, \bar{y}$) you get. Since prediction lies beyond the sampled data, it is as far from the mean as you can get, and thus the errors are compounded at their largest state.

The safety analysis for the Fukushima Daiichi nuclear power plant was based on historical data dating back to 1600 CE. The plant was designed to withstand a maximum earthquake of 8.6 magnitude, and a tsunami as high as 5.7 meters. The earthquake on March 11, 2011 measured 9.0 and resulted in a >14 meter high tsunami. The design basis was developed from a mistake in the regression analysis of the historical earthquake data. The structural engineers responsible for the design overfit their model to the existing data, rather than use the accepted Gutenberg-Richter model, they saw a kink in the data and assumed that the appropriate regression was not linear, but rather a polynomial, resulting in the following:

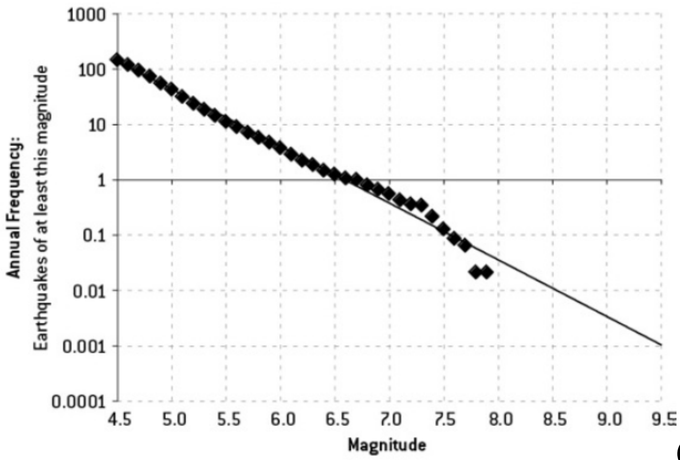
FIGURE 5-7C: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
CHARACTERISTIC FIT



(Silver, N, 2012)

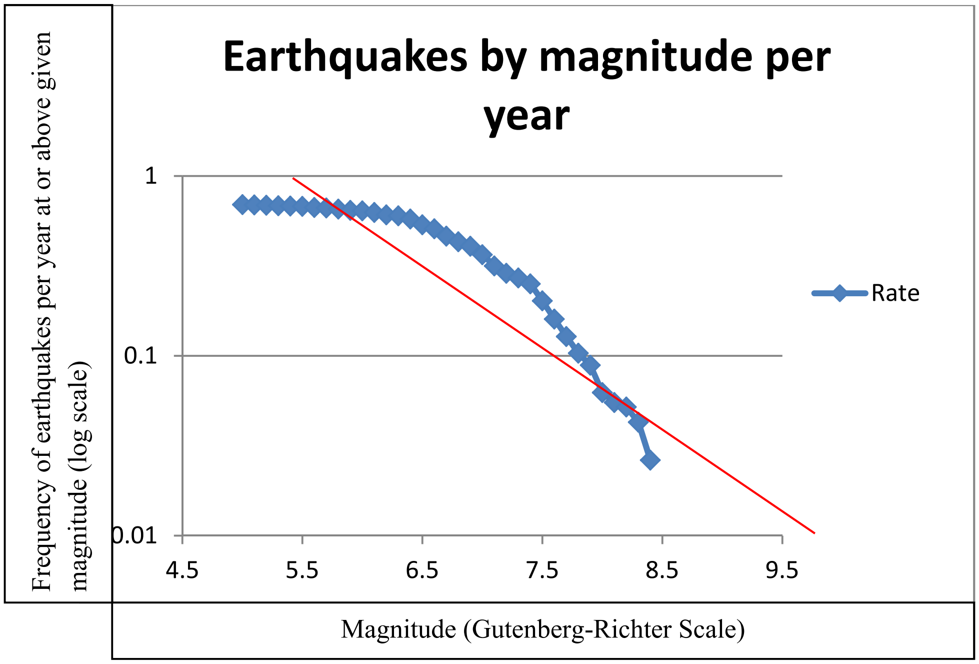
The Gutenberg-Richter method uses a simple linear regression to predict frequency versus magnitude based on historical data for a region and has been proven correct. That model for the same region looks like:

FIGURE 5-7B: TŌHOKU, JAPAN EARTHQUAKE FREQUENCIES
GUTENBERG-RICHTER FIT



(Silver, N, 2012)

This model clearly predicts that higher magnitude earthquakes are possible, and that a magnitude 9 earthquake is about 70 times more likely than predicted ($7.1x10^{-3}$ versus $1x10^{-4}$). The data I was able to obtain suggested that higher magnitude earthquakes would be even more likely. Data from the USGS concerning earthquakes in Japan since 1600 was used to develop the below graph. Plotted on a log-linear scale with a trendline added to show the x intercept (the Gutenberg-Richter model), this data shows that a magnitude 9.0 earthquake maybe as likely as $5x10^{-2}$. Based on historical observations, it is unlikely that my model is correct; however it does illustrate how sensitive these types of models are to minor variation in data.



http://earthquake.usgs.gov/earthquakes/search/

The non-conservative assumption made by these engineers resulted in a design that was not able to withstand neither a strong enough earthquake, nor a high enough wave. The damage that resulted from this event is widely known. The long lasting effects on the Japanese and World economies are not fully known, however the cost per megawatt-hour to produce power in Japan increased in the wake of the event due to the shutting down of all nuclear plant in the country and subsequent slow rate of restart by approximately $29.50, resulting in additional fossil fuel imports to the tune of $75.6 billion ("Reactor restarts would aid Japanese recovery", 2014). This apparently small assumption has resulted in a hug impact to Japan's economy, and should be a lesson to all on the power of predictive models.

References

Flom, P. (2011, June 8). The Disadvantages of Linear Regression. Retrieved March 29, 2015, from http://www.ehow.com/info_8562780_disadvantages-linear-regression.html

Green, S. (1991). How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behavioral Research, 26*(3), 499-510. Retrieved March 29, 2015, from http://www.tandfonline.com/doi/abs/10.1207/s15327906mbr2603_7#.VRg0GPnN8mk

Kamer-Ainur, A. (2007, January 1). Errors and Limitations Associated With Regression and Correlation Analysis. Retrieved March 29, 2015, from http://steconomiceuoradea.ro/anale/volume/2007/v2-statistics-and-economic-informatics/1.pdf

Ragheb, M. (2012, March 28). Fukushima Earthquake and Tsunami Station Blackout Accident. Retrieved March 31, 2015, from http://mragheb.com/NPRE 457 CSE 462 Safety Analysis of Nuclear Reactor Systems/Fukushima Earthquake and Tsunami Station Blackout Accident.pdf

Reactor restarts would aid Japanese recovery. (2014, July 28). Retrieved April 1, 2015.

Silver, N. (2012). *The signal and the noise: Why so many predictions fail--but some don't*. New York, NY: Penguin Press.

Verma, M., Chahal, G., & Murty, B. (1978). Limitations of conventional regression analysis a proposed modification. *Theoretical and Applied Genetics, 53*(2), 89-91. Retrieved March 29, 2015, from http://link.springer.com/article/10.1007/BF00817837