Econometric Forecasting Overview

April 30, 2014

Econometric Forecasting

- Econometric models attempt to quantify the relationship between the parameter of interest (dependent variable) and a number of factors (explanatory variables) that affect the dependent variable.
- Example
  - Dependent variable
    - Electric energy
  - Explanatory variable
    - Economic activity
    - Weather (HDD/CDD)
    - Electricity price
    - Natural gas price

Estimating Relationships

- Each explanatory variable affects the output variable in a different way. The relationships (or sensitivities) can be calculated via a number of different methods
  - Can be linear, polynomial, logarithmic, …

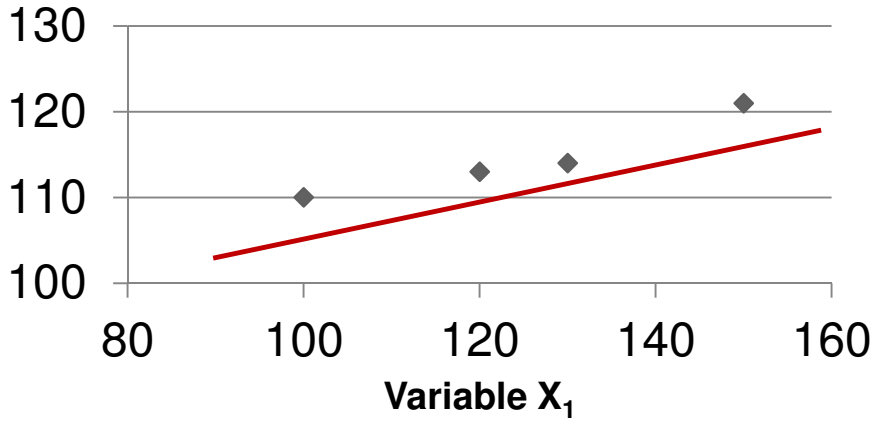
$$Y = \beta + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \ldots$$

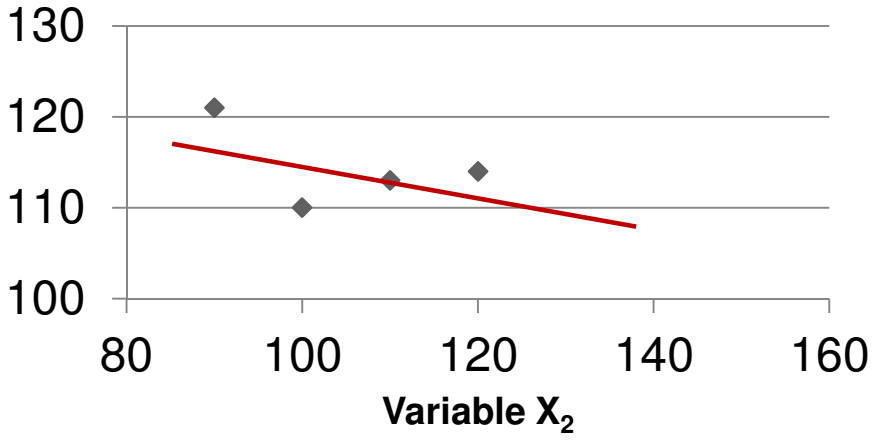
- Relationships are determined simultaneously to find overall best fit

A Simple Example

- Suppose we have 4 sets of observations with 2 possible explanatory variables

Output Y	Variable X_1	Variable X_2
110	100	100
113	120	110
114	130	90
121	150	120





A Simple Example

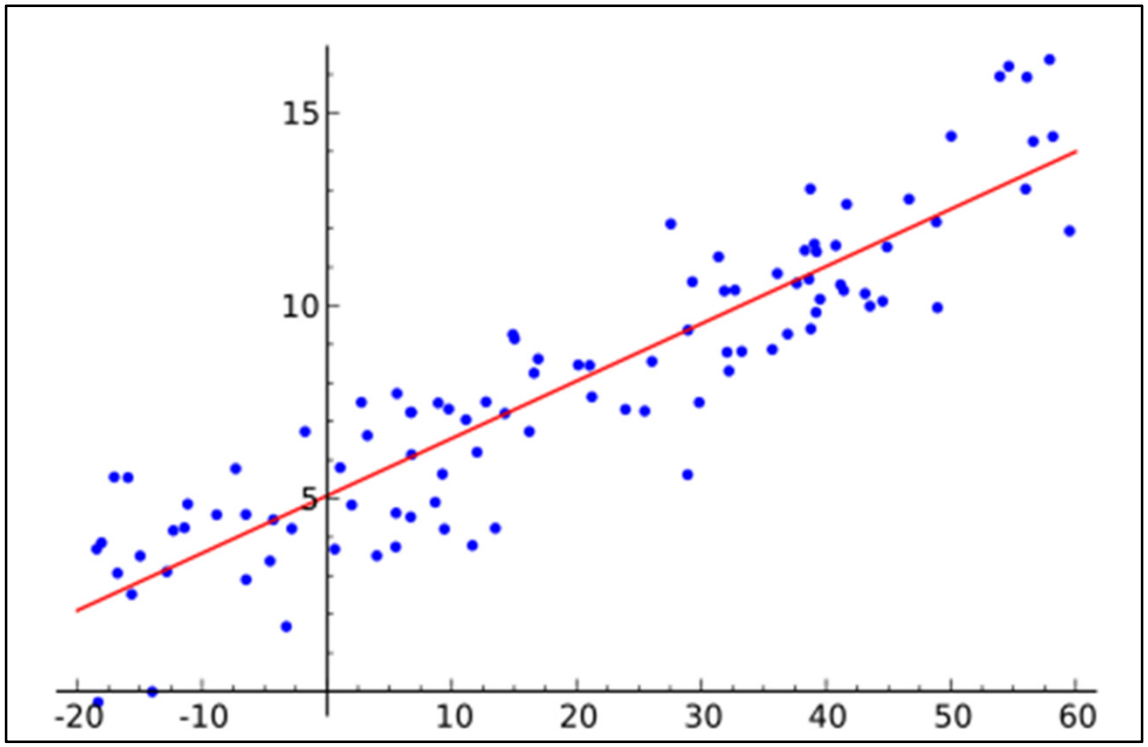
- Including both variables provides a perfect fit
  - Perfect fits are not usually achievable in complex systems

$$Y = 0.2X_1 - 0.1X_2 + 100$$

Finding a Good Fit

- A number of methods exist for finding the best fitting line, the most common is the ordinary least squares method
  - Ordinary least squares: find the line that minimizes the sum of the squares of the differences between the historical observations and the line
- Often, the “best” fit may not be desirable due to other statistical considerations
  - Some judgment is required to balance various factors

Regression Analysis



Source: Wikipedia.org

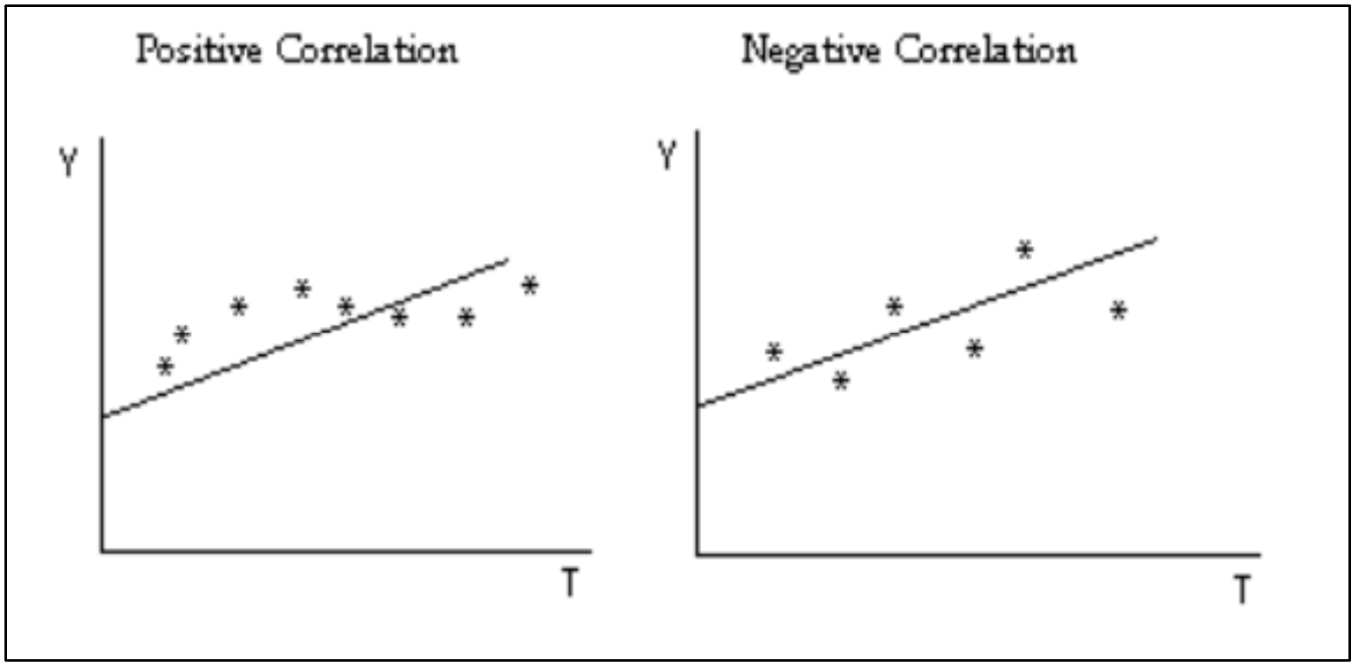
Model Error

- In simple terms, model error is the difference between the actual observed dependent variable value and the value that would be derived from the model equation
- Forecasters try to keep model error low but some error is generally unavoidable

Serial Correlation & Heteroskedasticity

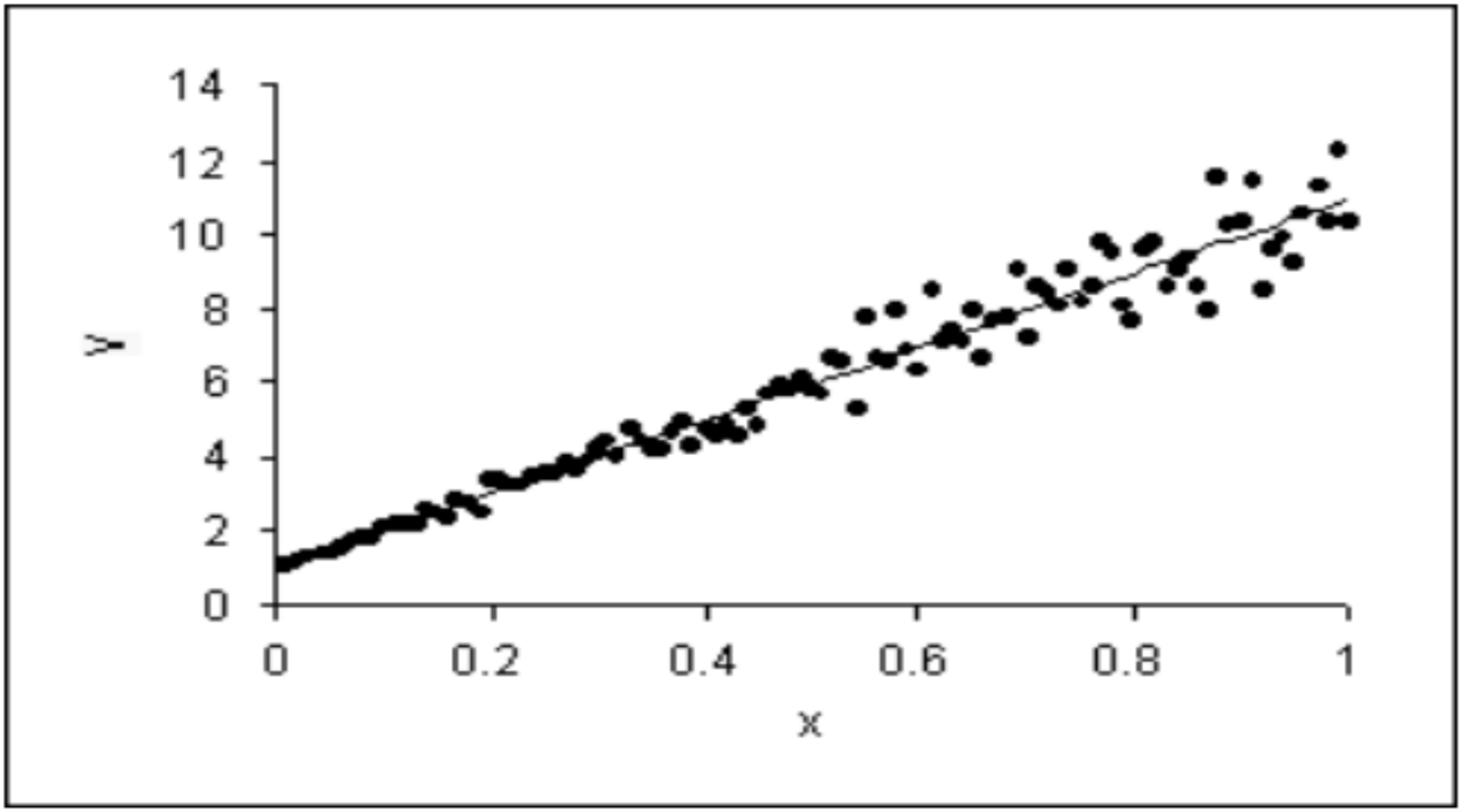
- Serial correlation occurs when the error terms from different time periods are correlated
- Heteroskedasticity occurs when the error terms have different variances
- These do not bias the model but results in a false level of accuracy
  - The model estimates may appear to be more precise than they actually are

Serial Correlation



Source: Limitations of the Classical Model, University of Rhode Island
http://www.uri.edu/artsci/newecn/Classes/Art/306a/Outlines/Statistics/Out.Stats2e.html

Heteroskedasticity



Heteroskedasticity in a simple, bivariate model.

Source: Introductory Econometrics, Cambridge University Press, 2005

Goodness of Fit

- R^2 – the proportion of the variation in the dependent variable that is explained by the independent variables
- Adjusted R^2 – a similar measure that penalizes for the inclusion of additional explanatory variables
- F-Statistic – a significant F-test indicates that the R^2 is reliable and not a spurious result.

Estimating a Model

- Specify the explanatory variables to be used
- Software (EViews) will determine estimation and provide statistical information
- Repeat using other sets of explanatory variables

Possible Model Formulations

- There are a number of different formulations that can be considered
  - Linear vs. logarithmic
  - First differences (using the change in value between adjacent time periods instead of the values themselves)
  - Moving average or lagged prices
  - Adjust the number of historical observations to be used

Choosing a Model

- We looked for models that provide a good fit (R^2 and Adjusted R^2 close to 1 and F-Statistic probability less than 0.05), pass the tests for serial correlation and heteroskedasticity, and contain an appropriate mix of explanatory variables with correct estimated coefficient signs.

Statistical Tests

- The Durbin-Watson statistic is a check for first-order serial correlation. The statistic ranges from 0 to 4 with 2 meaning no serial correlation; less than 2 being evidence of positive serial correlation; and greater than 2 being evidence of negative serial correlation.

Statistical Tests

- The Correlogram of Residuals and the associated Ljung-Box Q-Statistics up to 12 lags were also inspected to make sure there was no statistically significant serial correlation. (Q-Statistic associated probabilities are greater than 0.05)

Statistical Tests

- The White test was used to check for heteroskedasticity. The test statistic's associated probabilities should be greater than 0.05.