

Outliers

Problem

Regression parameters can be influenced by a few extreme values or outliers

- Should be able to spot from a careful analysis of the residuals \hat{u}_t
- In the case of a simple bivariate regression you can simply plot the data.
- Outlier is an observation that is very different: usually generated by some unusual factor
- Least squares estimates are very sensitive to outliers, particularly in small samples
- Maddala P89-90 gives examples of data sets that when plotted look very different, but give the same regression results, but look quite different. In two cases this is caused by a single extreme value.

Actions

- Drop the observations with large residuals and reestimate the equation. This should really be a last resort
- The outliers may provide important information. They may not be outliers at all. An example of this is the relation between infant mortality and GDP per capita in Asian countries.
- For cross section should maybe try to get more data rather than drop observations. Also for time series.
- Problem of what is an outlier also relates to leverage: need variation in the data or cant estimate any relationship. It's not always obvious when information on a system becomes an outlier.
- Can treat extreme observations with dummy variables