Seeing the "Man Who isn't There"

Find the sample observations with the largest positive residuals, and those with the largest (in magnitude) negative residuals. If some as-yet-not-in-your-model factor seems to differentiate the two groups, collect data on that factor and try including it as a new explanatory variable in your model.

Interactions

When ***the* effect (i.e., the coefficient) of** one explanatory variable ***on*** the dependent variable ***depends on*** the value of another explanatory variable, introduce an artificial product variable.

- Signaled only by judgment
- The "trick": Introduce the product of the two explanatory variables as a new artificial explanatory variable. After the regression, interpret in the original "conceptual" model.
- For example, Cost = a + (b_1+b_2·Age)·Mileage + … (rest of model)
- The latter explanatory variable (in the example, Age) might or might not remain in the model
- Cost: We lose a meaningful interpretation of the beta-weights

Nonlinearities

When the direct relationship between an explanatory variable and the dependent variable "bends" (signaled by a "U" in a plot of the residuals against an explanatory variable), introduce the square of that variable as a new artificial explanatory variable: $Y = a + bX + cX^2 + \ldots$ (rest of model)

- This one "trick" can capture 6 different nonlinear "shapes".
- Always keep the original variable (the linear term, with coefficient "b", allows the parabola to take any horizontal position).
- The sign of c tells you the orientation of the fitted parabola (positive = upward-bending parabola, negative = downward-bending).
- -b/(2c) indicates the value of x where the vertex (either maximum or minimum) of the parabola occurs.
- Cost: We lose a meaningful interpretation of the beta-weights.