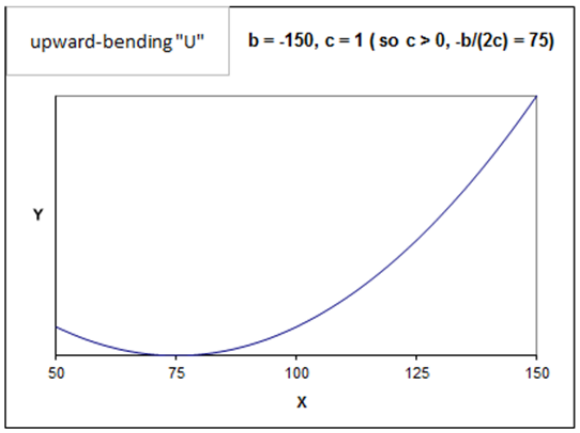
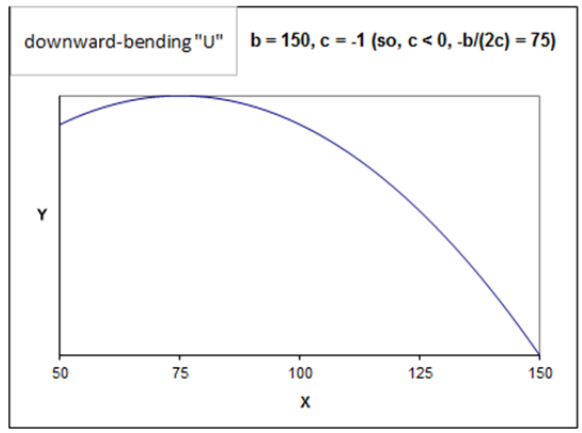
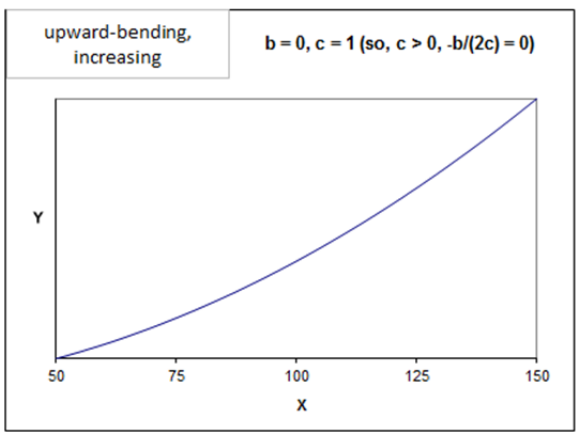
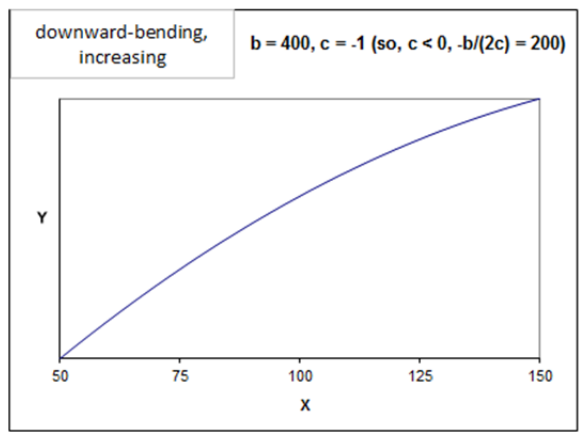
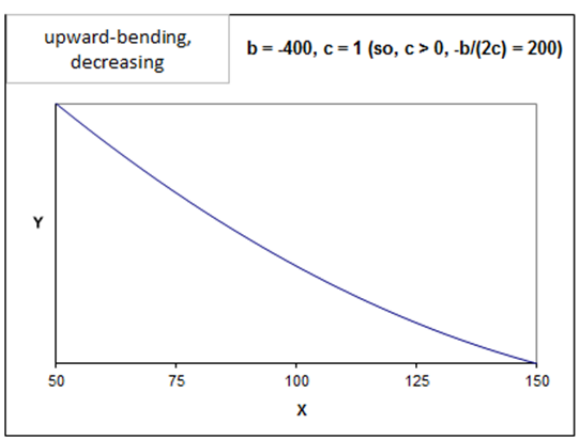
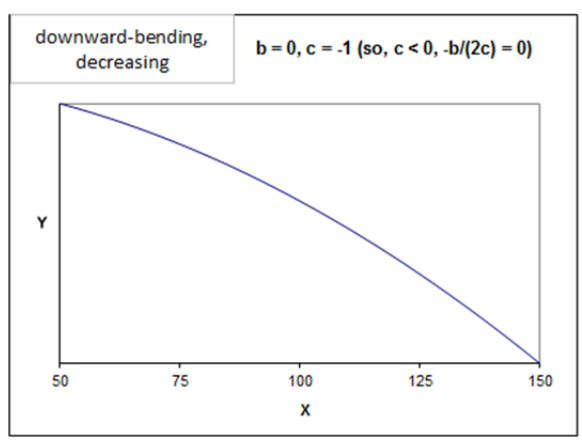
Capturing Nonlinearity

In order to capture a curved (i.e., nonlinear) relationship, we can include the square of an explanatory variable (together with the original variable) in our regression model. The model now takes the form: $Y = \ldots + bX + cX^2 + \ldots$.

The values of the two coefficients together determine the shape of the curve. If $c > 0$, the curve bends upwards (and the magnitude of c determines the "tightness" of the bend); if $c < 0$, the curve bends downwards.

The curve will be a portion of a parabola which takes its lowest ($c > 0$) or highest ($c < 0$) value when $X = -b/(2c)$. If this value of X is within the range of observed values of X, we'll actually see some of both "wings" of the parabola in the captured relationship. However, if this value is outside our observed data range, only part of one wing will actually appear.

In the examples below, the range of observed values of X is assumed to be between 50 and 150.