Quadratics and Higher-Order Polynomials

Sometimes we want a variable to be entered in our regression as a polynomial.

- Guiding theory requires it.
- Previous data suggests that relationship is nonlinear.
- We want to allow effect of x on y to change with value of x.
- We're looking to reduce error in prediction by allowing more flexibility in the functional form.
  - Especially for control variables

Guidelines

- For quadratic form, OLS will find parabola with least squared error; for cubic form, OLS will find best cubic function.
  - Fit improves when adding higher-order term.
- Include all lower-order terms (if x^2, then x too).
- Interpret all coefficients simultaneously to understand effect of variable.

Quadratic Form Example

Fitted wage equation, using experience and experience-squared as predictors.

$$\underset{}{\hat{wage}} = \underset{(.35)}{3.73} + \underset{(.041)}{.298}exper - \underset{(.0009)}{.0061}exper^2$$

$n = 526,\ \mathrm{R}^2 = .093$

- Coefficient for exper^2 is negative; function is **concave**.
- Every extra year of experience adds less to wage than previous year.

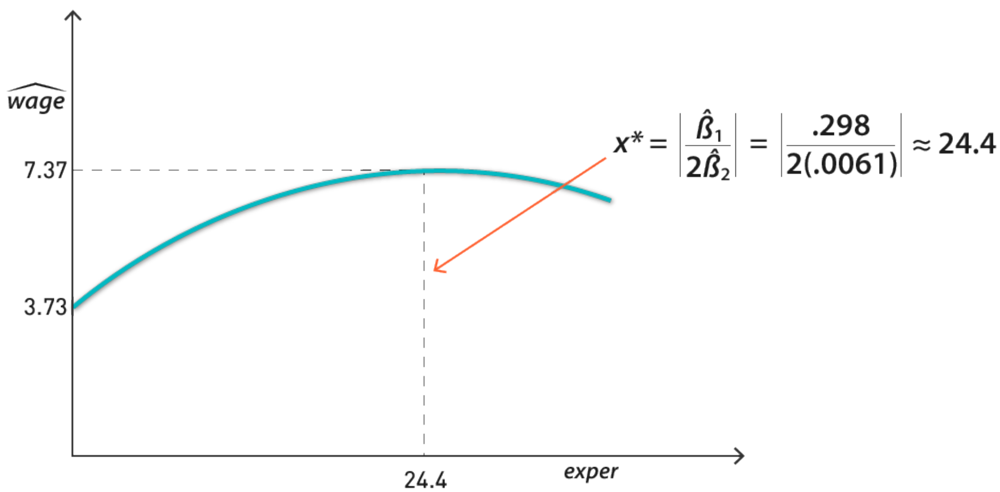
Quadratic Form Example (cont.)

Take derivative of both sides:

$$\frac{\delta wage}{\delta exper} = .298 - 2\left(.0061\right)exper$$

- Slope is decreasing.
- First year of experience increases wage by about \$0.30, the second year by .298 – 2(.0061)(1) = \$0.29, etc.
- Experience has decreasing marginal benefit.

Fitted Quadratic Wage Model



Extrapolating to Extreme Values

Does this mean that wage will start falling after 24.4 years?

- Not necessarily:
  - There may not be many data points above 24.4 years.
  - We don't know if model specification is correct.
    - Any concave parabola will have negative slope after some point.

We need to be careful when extrapolating the results of a model to extreme values.