

# Regression User's Guide (1 of 2)

| What Can Go Wrong?   | What Are the Consequences?   | How Can It Be Detected?  | How Can It Be Corrected?   |
|--|--|--|--|
| <b>Omitted Variable</b>  |  |  |  |
| The omission of a relevant independent variable                  | Bias in the coefficient estimates (the $\hat{\beta}$ s) of the included Xs.                                  | Theory, significant unexpected signs, or surprisingly poor fits.   | Include the omitted variable or a proxy.                                       |
| <b>Irrelevant Variable</b>                                       |  |  |  |
| The inclusion of a variable that does not belong in the equation | Decreased precision in the form of higher standard errors, lower $t$ -scores and wider confidence intervals. | <ol style="list-style-type: none"> <li>1. Theory</li> <li>2. <math>t</math>-test on <math>\hat{\beta}</math></li> <li>3. <math>\bar{R}^2</math></li> <li>4. Impact on other coefficients if X is dropped.</li> </ol> | Delete the variable if its inclusion is not required by the underlying theory. |
| <b>Incorrect Functional Form</b>                                 |  |  |  |
| The functional form is inappropriate                             | Biased estimates, poor fit, and difficult interpretation.  | Examine the theory carefully; think about the relationship between X and Y.  | Transform the variable or the equation to a different functional form.         |

# Regression User's Guide (2 of 2)

| What Can Go Wrong?   | What Are the Consequences?   | How Can It Be Detected?               | How Can It Be Corrected?  |
|--|--|---------------------------------------|---|
| <b>Multicollinearity</b><br>Some of the independent variables are (imperfectly) correlated   | No biased $\hat{\beta}$ s, but estimates of the separate effects of the Xs are not reliable, i.e., high $SE(\hat{\beta})$ s and low $t$ -scores. | Pairwise correlations or scatterplots | Drop redundant variables, but to drop others might introduce bias. Often doing nothing is best.   |
| <b>Serial Correlation</b><br>Observations of the error term are correlated, as in:<br>$\epsilon_t = \rho\epsilon_{t-1} + u_t$                | No biased $\hat{\beta}$ s, but OLS no longer is minimum variance, and hypothesis testing and confidence intervals are unreliable.                | Use residual plots                    | If impure, fix the specification.   |
| <b>Heteroskedasticity</b><br>The variance of the error term is not constant for all observations, as in:<br>$VAR(\epsilon_i) = \sigma^2 Z_i$ | Same as for serial correlation.  | Use residual plots                    | If impure, fix the specification. Otherwise, use robust std. errors or reformulate the variables. |