Regression Modeling Best Practices

1. Clean the data before you begin modeling. Make sure units are consistent and data is as error free as possible.
2. Graph the data and perform visualizations to look at the structure of the data. This will help you determine whether to use linear regression, nonlinear regression, or some other technique.
3. Keep models simple – use a minimum number of explanatory variables.
4. If you have enough data, hold out some data in a test set. If you have a smaller data set, use cross validation. This will help you avoid overfitting.
5. Avoid stepwise regression for variable selection unless you use a test set or cross-validation to avoid overfitting.
6. Avoid bias in your models. For nonlinear methods, the use of log-transformed least squares underestimates the mean. Avoid using log transforms or correct for their bias.
7. Examine your model residuals. Some methods require specific residual distributions to be optimal, such as they need to follow a normal or lognormal distribution.