DIAGNOSING MODEL PROBLEMS

**Multicollinearity** - a problem in which two independent variables in a model are highly related. Multicollinearity is a sample phenomenon, not an estimation problem.

NOTE: If all of the independent variables are checked, and the t-tests are significant, then multicollinearity should not be a problem.

**PROBLEMS caused by Multicollinearity:**
1) The estimated coefficient ($\beta$) may no longer be an unbiased estimator.
2) Because of the strong interrelationships between the independent variables, it is difficult to disentangle their separate effects on the dependent variable without violating the *ceteris paribus* assumption.
3) Significant variables may appear to be insignificant. (i.e., a Type II error)

**CHECKING for Multicollinearity:**
1) High $R^2$ with few or no significant t-ratios.
2) Variance Inflation Factor (VIF)
   If $VIF > 10$, then suspect a problem.
   $VIF = 1$ is IDEAL!!!
3A) & 3B) should be performed jointly.
3A) Analysis of Structure (SAS Procedure)
3B) Condition Index (CI)
   If $CI > 30$, then suspect a problem.
4) Variance Proportion
   If two or more Variance Proportions > 0.5 in a single row, then suspect a problem.
5) High pairwise correlations among independent variables.

**POSSIBLE SOLUTIONS to Multicollinearity:**
1) Dropping the redundant independent variable; however, be careful not to omit a relevant variable.* <See Multicollinearity>
2) Combining highly correlated variables
   (NOTE: use of interaction variables.)
3) Using ratios on a first difference; however this may introduce a Heteroscedasticity problem.
4) Using ridge regression.
5) Using principal component analysis.
6) Getting more data.
7) Do Nothing! Live with the lesser of two evils.*<See Multicollinearity definition for warning>

**Heteroscedasticity** - a problem in which the variance of the residuals along a trend line is not constant.

NOTE: The coefficient ($\beta$) remains an unbiased estimator nonetheless.

**PROBLEMS caused by Heteroscedasticity:**
1) The estimated coefficient ($\beta$) is not "efficient" or the "best" estimator.
2) The F-test is no longer reliable.
3) The t-tests are no longer reliable.

**CHECKING for Heteroscedasticity:**
1) Check the scatterplot of residuals for fan shaped patterns.
2) Goldfeld-Quandt test
3) Breusch-Pagan test
4) White test

**POSSIBLE SOLUTIONS to Heteroscedasticity:**
1) Generalized Least Squares
2) Weighted Least Squares
3) Use weighting via generalized differencing if the true variance is known.
4) Perform a transformation of the model from linear to log-linear form.

**Autocorrelation** - a problem in which the residuals are not independent. This problem is often seen in time-series data. It may be the result of an omitted relevant variable* or an incorrect functional form.

NOTE: may be positively or negatively related.
NOTE: the coefficient ($\beta$) remains an unbiased estimator nonetheless.

**caused by Autocorrelation:**
1) The variance of the coefficient ($\beta$) may be understated.
2) The F-test is no longer reliable.
3) The t-tests are no longer reliable.

**for Autocorrelation:**
1) Check the scatterplot of residuals for signs of patterns. (positive or negative autocorrelation)
2) Durbin-Watson statistic
to Autocorrelation:
1) Generalized differencing (if $p$ is known)
2) The Cochran-Orcutt procedure
3) The Hildreth-Lu procedure
4) Durbin's H test - USE only when there is a lagged dependent variable.

© Roger E. Wehr 1999