## Coping With Collinearity

**Model Respecification**

In some cases collinearity can be addressed by combining covariates or eliminating some covariates from a set of highly correlated covariates.

**Variable Selection**

Another approach for lessening the effects of collinearity is by using a variable selection procedure. Older procedures such as forward selection, backward selection and stepwise selection place too much emphasis on a single final model, but more recent approaches to all-subset selection methods allow comparison of many models. Typically there will be several models that are close to the optimum according to a chosen criterion, illustrating that several candidate models may be useful. Several of the optimality criteria for model selection are discussed in Chapter 22 including the class of information criteria, which assigns a score to the $j^{th}$ model of

$$IC_j = -2\log L(\widehat{\theta}_j) + cs_j,$$

where $\log L(\widehat{\theta}_j)$ is the maximized log-likelihood value, $s_j$ is the number of parameters in $\theta_j$, and $c$ is a constant that differs between different $IC$ criteria. The two most well-known criteria are the Akaike Information criterion (AIC) and the Bayesian Information Criterion (BIC), defined as:

$$\begin{aligned} AIC_j &= -2\log L(\widehat{\theta}_j) + 2s_j, \text{ and} \\ BIC_j &= -2\log L(\widehat{\theta}_j) + s_j \log(n). \end{aligned}$$

For the linear model with normal errors, these expressions simplify to

$$\begin{aligned} AIC_j &= n\log \widehat{\sigma}_\varepsilon^{(j)2} + 2s_j, \text{ and} \\ BIC_j &= n\log \widehat{\sigma}_\varepsilon^{(j)2} + s_j \log(n). \end{aligned}$$

The model with the smallest value is best for either IC criterion.

Another commonly-used model selection criterion is Mallow's $C_p$ statistic, which attempts to balance the tradeoff between bias (typically due to underfitting) and variance (typically due to overfitting) in the mean-squared error of predictions from a regression model. Mallow's $C_p$ statistic is

$$C_{pj} = \frac{\sum E_i^{(j)2}}{S_E^2} + 2s_j - n,$$

where $E_i^{(j)}$ are the residuals from model $M_j$, and $S_E^2$ is the error variance estimate from the full model with $k+1$ regressors. Under the null hypothesis that all regressors not in the $j^{th}$ model have coefficients equal to zero we have $E(C_{pj}) \approx s_j$, so the preferred model should have $C_{pj}$ approximately equal to or less then $s_j$.