Model Refinement : Added Variable Plots

The aim of multiple regression (MR) modelling is to use the X or explanatory variables / predictors to explain what is happening with the Y or response variable.

In analysis of variance terms we want to maximise the variance explained by the model (involving the X variables) & minimise the unexplained variance (in Y). In fact, we would like (and we assume) that this unexplained variance will be just random stochastic variability

ie $\underset{\sim}{\varepsilon} \overset{iid}{\sim} N(0, \sigma^2)$

"the errors are independent & identically (Normally) distributed with constant variance, σ^2"

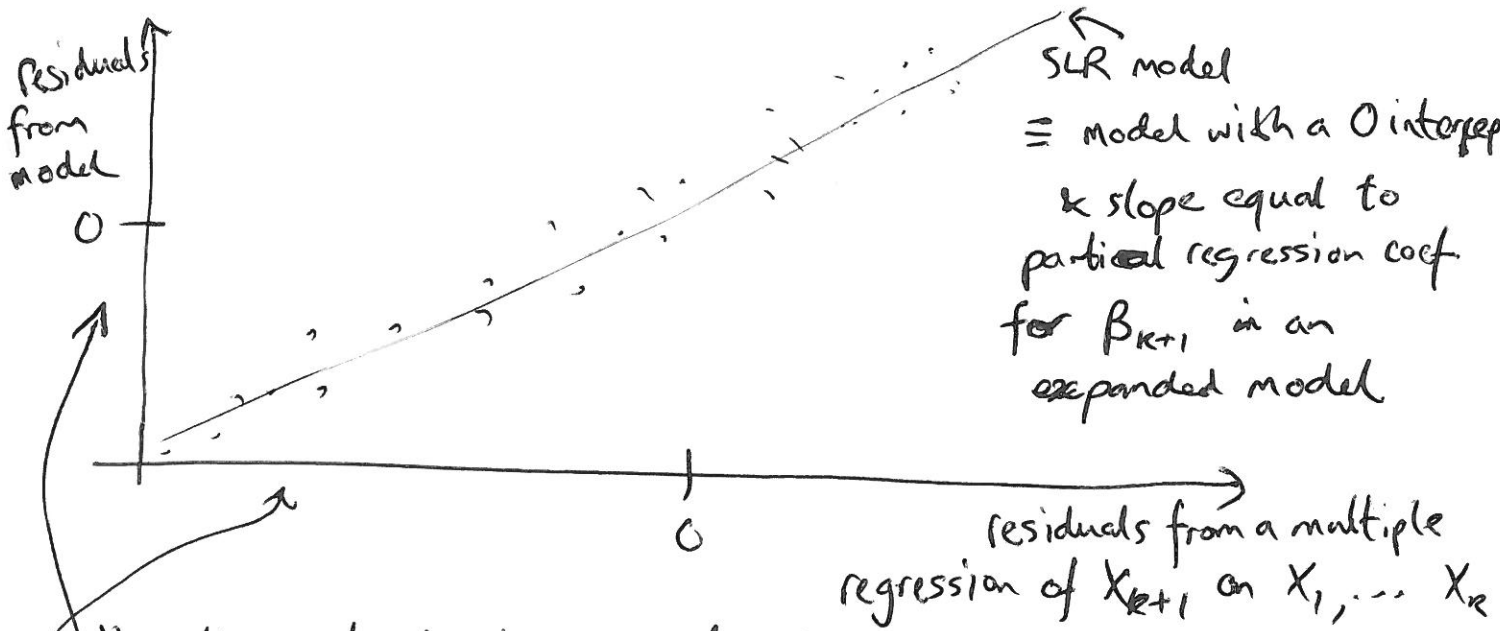
However, the residuals (observed errors) from a multiple regression model often show non-random structure (ie they are not independent) or they might be random, but simply too variable (the unexplained variance is large & the explained variance is relatively small & the overall F-test is not significant).

If we have other X variables (in the data, but not in the model), we can try adding them to the model & see if they help explain the unwanted structure and/or some of the unexplained variation.

Added variable plots (AVP) are a way of assessing if, but also how, we might include another X variable in the model.

Note, if we don't have any other candidate X variables in the data (there are no other observed X variables) then the structure and/or excess variability is "unobserved heterogeneity"

AVP for a model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + \varepsilon$
and an additional variable X_{k+1}



Here the vertical axis represents what is left in Y that is not explained by $X_1, \ldots X_k$

& the horizontal axis is what is in X_{k+1}, that is not explained by $X_1, \ldots X_k$

We can check if the relationship on the AVP is linear by adding a SLR line

OR we could propose a different way to include X_{k+1} in the model, ie some transformation or higher order term. If there is an apparent relationship which is non-linear!