Supplementary Material to Lesson 8-9: ECONOMETRICS

Generalized Regression and ANOVA models:

1. *Heteroscedascity*: $\mathbf{Y} = \mathbf{X}\mathbf{a} + \underline{\varepsilon}$, $\mathbb{D}^2(\underline{\varepsilon}) = \sigma^2\mathbf{B}$, where $\mathbf{B} > 0$ is given, σ^2 and $\mathbf{a}$ are to be estimated. Solved by transforming the data: with $\mathbf{Z} := \mathbf{B}^{-1/2}\mathbf{Y}$, $\mathbf{U} := \mathbf{B}^{-1/2}\mathbf{X}$, the model is

$$\mathbf{Z} = \mathbf{U}\mathbf{a} + \underline{\varepsilon}', \quad \mathbb{D}^2(\underline{\varepsilon}') = \sigma^2\mathbf{I}.$$

   Hence, $\hat{\mathbf{a}} = (\mathbf{U}^T\mathbf{U})^-\mathbf{U}^T\mathbf{Z}$.

2. *Time series*: the data represents a series of observations taken over time, and covariances are *autocovariances* between the lagged observations. Solved by time series techniques.
3. *Dependence*: $Y = \mathbf{a}^T\mathbf{X} + \varepsilon$, where ε is correlated with $\mathbf{X}$. Solved by *instrumental variables*: $\mathbf{Z}$ is highly correlated with $\mathbf{X}$, but uncorrelated with the disturbance ε.
4. *Analysis of Covariance*: ANOVA model with *concomitant variables*:

$$\mathbb{E}(\mathbf{Y}) = \mathbf{B}\mathbf{a} + \mathbf{D}\mathbf{c},$$

   where $\mathbf{B}$ is $n \times k$ *design matrix* of entries 0-1, $\mathbf{D}$ is $n \times \ell$ matrix of observations on the (continuous) concomitant variables; further, $\mathbf{a} \in \mathbb{R}^k$ and $\mathbf{c} \in \mathbb{R}^\ell$ are parameters to be estimated by minimizing

$$\sum_{i=1}^{n}(Y_i - b_{i1}a_1 - \cdots - b_{ik}a_k - d_{i1}c_1 - \cdots - d_{i\ell}c_\ell)^2.$$

   The normal equasions form the following system:

$$\mathbf{B}^T\mathbf{B}\mathbf{a} + \mathbf{B}^T\mathbf{D}\mathbf{c} = \mathbf{B}^T\mathbf{Y},$$
$$\mathbf{D}^T\mathbf{B}\mathbf{a} + \mathbf{D}^T\mathbf{D}\mathbf{c} = \mathbf{D}^T\mathbf{Y}.$$

   If we can accept the zero hypothesis $\mathbf{c} = \mathbf{0}$, then we perform usual one-way ANOVA; otherwise, we eliminate the effect of the concomitant variables. See C. R. Rao: Linear statistical inference, page 291, where Y is the growth rate of pigs, the $k = 3$ factors are pen, sex, and type of food given, while the only concomitant variable is the initial weight.