CRG 17/18 Meeting 9
Confounding and Efficiency in Clustered Data (Vansteelandt,2007)

Discussant: Oisín Ryan

April 12, 2018

Background: GEE

- Typical multi-level (linear mixed effects) modeling scenario
  - (Nested) data modelled with a **gaussian** distribution
  - Derive a likelihood for the data and maximise
    - **Score function**=0
  - Fixed effects - population average parameter
  - Fixed effects also referred to as the **Marginal** part
  - Random part - individual-specific difference from the marginal parameter
  - Estimate the variance of the random part

Background:GEE

- Problem: not straightforward for non-gaussian distributions
  - GLM without identity link, e.g. Bernoulli, Poisson etc.
  - Joint likelihood difficult to specify
  - Problem arises from complex variance function - due to the random part

- Solution: **Generalized Estimating Equations** (GEE)
  - Only care about estimating the **Marginal Part** (mean response)
  - Treat the random effects part as a nuisance parameter
  - Semi-parametric: Don't have to specify the full likelihood, only first moment
  - Get a similar looking **score function**
  - Upshot: marginal parameters only dependent on the first moment, so we can mis-specify the variance/covariance structure and still get good estimates! (black magic)
    - Misspecified variance/covariance = Working correlation structure
  - Downshot: No nice estimates of random part, SEs need to be corrected later

Conditional Mean Model for Longitudinal Data

$$E(Y_t|\bar{X}_t) = h_t(\bar{X}_t; \omega*) \tag{1}$$

- $\bar{X}_t$ is potentially all values of the predictor variable at all points in time up to and including t
- ω are the parameters relating $\bar{X}_t$ to Y_t

$$\begin{aligned} E(Y_t|\bar{X}_t) &= \omega_0 + \omega_1 X_{t-1} \\ &= \omega_0 + \omega_1 X_t + \omega_2 X_{t-1} \\ &= \omega_0 + \omega_1 t + \omega_2 X_t + \omega_3 X_{t-1} \end{aligned}$$

The solution to the usual GEEs

$$\sum_{i=1}^{n} \Gamma_i \Sigma_i^{-1} \epsilon_i(\omega) = 0 \tag{2}$$

This is just the **score function** where

- $\epsilon_i(\omega)$ is the error
- Σ_i is the variance covariance matrix of the errors
- Γ_i are the derivatives of the predictor equation with respect to the parameters

Compare to the score function of a gaussian GLM

$$S(\boldsymbol{\beta}) = \sum_i \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} v_i^{-1} (y_i - \mu_i) = 0$$

Unbiasedness conditions

GEE Estimates only guaranteed to be unbiased when

$$E(Y_t|\bar{X}_t) = E(Y_t|\bar{X}_T) \tag{3}$$

Current values of Y are independent of future values of X given current values of X