Lecture 20

Find Wed 10.30 - 12.30
Economic Blue Conference Room

Q1: Basic statistical learning (answer 5 of 7 questions)
Q2: Bayesian (no multiple inputs)
Q3: Clustering
Clustering

\[ \hat{\beta} = (X'X)^{-1} X'y \]

\[ = \beta + (X'X)^{-1} X'\epsilon \]

\[ \text{Var}(\hat{\beta}) = (X'X)^{-1} \text{Var}(X'\epsilon) (X'X)^{-1} \]

\[ = (X'X)^{-1} E\epsilon \epsilon' (X'X)^{-1} \]

\[ \text{Var}(\hat{\beta}) = (X'X)^{-1} X'\hat{\epsilon} \hat{\epsilon}' X (X'X)^{-1} \]

Does not work since \( X'\hat{\epsilon} = 0 \). By 

by our f.o.c., cluster robust case \( E\epsilon \epsilon' \) is block diagonal

\[ = (X'X)^{-1} \sum_{g} X_g' \hat{\epsilon}_g \hat{\epsilon}_g' X_g (X'X)^{-1}. \]
Extends to Feasible GLS

GLS \[ \hat{\beta} = (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y \]

\[ = \beta + (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} \nu \]

\[ \text{Var} (\hat{\beta}) = (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} E(\nu \nu' | X) \Lambda^{-1} X (X' \Lambda^{-1} X)^{-1} \]

\[ \text{Var} (\hat{\beta}_{FGLS}) = (X' \Lambda^{-1} X)^{-1} \sum_{g} x_{g} \Lambda^{-1} \hat{\nu}_{g} \hat{\nu}_{g} \Lambda^{-1} x_{g} (X' \Lambda^{-1} X)^{-1} \]

Given clustering \( y \) so \( \Lambda = \begin{bmatrix} \Lambda_{1} & 0 \\ 0 & \Lambda_{G} \end{bmatrix} \)

when \( \hat{\nu}_{g} = y_{g} - x_{g} \hat{\beta}_{FGLS} \)
- Potential efficiency gains for FGLS
  Have a reasonable model for \( \pi \) (working matrix)
  but do robust inference that does not require correct \( \pi \)
  (do need true \( \pi \) is block diagonal and \( G \to \infty \))

- For individuals clustered in region will age, ...
  RE model \( \sigma_g = \sigma^2 \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \)
  or richer mixed models with random slopes
  not just random intercept

Big deal everywhere but econometrics.
Hierarchical linear models. State mixed, VCE(robust)
cluster.
For peel deck

AR(1) \( u_{it} = \rho u_{it-1} + \varepsilon_{it} \)

\[ \Lambda_g = \sigma_u^2 \left[ \begin{array}{cccc}
1 & \rho & \ldots & \rho^{n_g-1} \\
\rho & \ldots & \rho \\
& \ldots & & \\
\rho & & & 1
\end{array} \right] \]

MA(1) \( \Lambda_g = \sigma_u^2 \left[ \begin{array}{cccc}
\rho & \rho & & \\
& \rho & \ldots & \\
& & \rho & \\
& & & \rho
\end{array} \right] \)
- Cluster-robust generalizes to pretty much all estimators. For consistency we need to assume that clustering does not lead to the normal condition of the estimator no longer holding.

- m-estimators

\[ \mathbb{E} \sum m(y_i, x_i, \Theta) = 0 \]
\[ \text{logit} \quad \mathbb{E} \sum (y_i \log \Lambda(x_i^\prime \beta)) \cdot x_i = 0 \]
\[ \Delta(\tau) = \frac{e^\tau}{1 + e^\tau} \]
\[ \text{i}(0,1) \]

since

\[ \log L(\beta) = \sum \log \left( \Lambda(x_i^\prime \beta)^{y_i} \times (1 - \Lambda(x_i^\prime \beta))^{1-y_i} \right) \]

\[ \frac{\partial \log L(\beta)}{\partial \beta} = \sum \left( y_i \cdot \Lambda(x_i^\prime \beta) \cdot x_i \right) = 0 \]
- Two approaches

**MLE. Bring in a model for clus"tering**

\[ P(y_i = 1) = \Lambda(x_i' \beta + \alpha_i) \quad \alpha_i \sim N(0, \sigma^2) \]

Assume correctly specified, no cluster robust.

This is \texttt{xtlogit}, re

- **Quasi-MLE. Population-averaged**

Assume \( P(y_i = 1) = \Lambda(x_i' \beta) \) even with correlated within cluster

But get cluster-robust s.e.

\[ \hat{\Theta} \text{ solves } \sum_g \sum_i m(y_{ig}, x_{ig}, \Theta) = 0 \]

\[ \sum_g m_g(\Theta) = 0 \quad m_g = \sum_i m(y_{ig}, x_{ig}, \Theta) \]
\[ \sum_g m_g(\theta) = 0 \]

Take first-order Taylor series expansion:

\[ \text{Var}(\hat{\theta}) = \left( \sum_g \frac{\partial m_g(\theta)}{\partial \theta} \right)' \left( \sum_g \text{Var}(m_g(\theta)) \right) \left( \sum_g \frac{\partial m_g(\theta)}{\partial \theta} \right)'^{-1} \]

\[ \text{sum} \quad \text{Var} \left( \sum_g m_g(\theta) \right) \]

\[ = \sum_g \text{Var}(m_g(\theta)) \quad \text{given independence across clusters} \]

\[ \text{Var}(\hat{\theta}) = \left( \sum_g \frac{\partial m_g(\theta)}{\partial \theta} \bigg|_{\tilde{\theta}} \right)' \left( \sum_g m_g(\theta)m_g(\theta)' \right) \left( \sum_g \frac{\partial m_g(\theta)}{\partial \theta} \bigg|_{\tilde{\theta}} \right)^{-1} . \]
- Rank deficiency

$$\sum_{g} x_g \hat{u}_g \hat{y}_g x_g = C' C \text{ is } k \times k$$

where \( C = \left[ X_1 \hat{u}_1 \cdots X_6 \hat{u}_6 \right] \) is \( \mathbb{R}^2 \times 6 \times k \)

\( C' C \) has rank at most \( \min(k, 6) \)

Also \( \sum_{g} x_g \hat{u}_g = 0 \) by OLS force

So rank \( N \) at most \( \min(k, 6) - 1 \).

Not unusual to have \( k = 25 \)

And have \( G = 20 \)

Cannot do the overall F-stat. But individual t's are okay