From CLRM to Multilevel Model

eusup_{ij} = \beta^0_j + \beta^{lr}_j \text{ftr}_{ij} + \ldots + \epsilon_{ij}

\beta^0_j = \alpha_0 + \alpha_1 \text{GSPEND}_j + u^0_j

\beta^{lr}_j = \gamma_0 + \gamma_1 \text{GSPEND}_j + u^1_j

- If CLRM assumptions apply, then OLS unbiased, consistent, and efficient.
  - Two main issues of concern:
    - Parameter heterogeneity: (see pictures)
      - systematic &/or stochastic (fixed v. rndm intrcpt/coeff)
      - can cause bias if pattern unmodeled hetero relates to X,
    - Non-spherical error cov-mat: an efficiency & proper s.e.’s issue, not a bias/consistency one
      - But “mere inefficiency” can be serious.
      - And accurate std err’s very important.
From the CLRM to HLM

- Examples of parameter heterogeneity that covaries w/ X values, so bias:

- Note: FE v. RE both theoretically could cause bias if cov w/ X, but latter i.d.’d off orthog
From the CLRM to RE Model

- **Std. R.E. Model**: Odd that std. lin-interax model:
  - Assumes know \( y = f(X) + \text{error} \):
    \[
y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_{xz} x_i z_i + \epsilon_i
    \]
  - But \( \frac{dy}{dx} = f(z) \) w/o error!:
    \[
    \frac{dy}{dx} = \beta_x + \beta_{xz} x z
    \]
  - So, try:
    \[
y = \beta_0 + \beta_1 x + \beta_2 z + \epsilon^0
    \]
    \[
    \frac{dy}{dx} \equiv \beta_1 = \alpha_0 + \alpha_1 z + \epsilon^1
    \]
    \[
    \frac{dy}{dz} \equiv \beta_2 = \gamma_0 + \gamma_1 x + \epsilon^2
    \]
    \[
    \Rightarrow y = \beta_0 + (\alpha_0 + \alpha_1 z + \epsilon^1) x + (\gamma_0 + \gamma_1 x + \epsilon^2) z + \epsilon^0
    \]
    \[
    = \beta_0 + \alpha_0 x + \gamma_0 z + (\alpha_1 + \gamma_1) x z + (\epsilon^0 + \epsilon^1 x + \epsilon^2 z)
    \]
  - => std. lin-interact...except compound error-term...

- **Std. HLM**: Same model, except \( x_{ij} \) & \( z_j \), &
  - So std lin-interact, but w/ diff compound-error struct.
    \[
    \epsilon^* = \epsilon_{ij}^0 + \epsilon_{ij}^1 x_{ij} + \epsilon_{ij}^2 z_j
    \]
From CLRM to Hierarchical Model

- Std. HLM: Same model, except $x_{ij}$ & $Z_j$ &
- So a std. lin-interact too, but with different compound-error stochastic properties.

\[
eusup_{ij} = \beta_0^0 + \beta_j^{lr} lft_{ij} + \ldots + \epsilon_{ij}
\]
\[
\beta_0^0 = \alpha_0 + \alpha_1 \text{GSPEND}_j + u_0^0
\]
\[
\beta_j^{lr} = \gamma_0 + \gamma_1 \text{GSPEND}_j + u_j^1
\]
\[\Rightarrow \quad \text{eusup}_{ij} = \alpha_0 + \alpha_1 \text{GSPEND}_j + u_j^0 + \gamma_0 lft_{ij}
\]
\[\quad + \gamma_1 lft_{ij} \times \text{GSPEND}_j + lft_{ij} u_j^1 + \ldots + \epsilon_{ij}
\]
gathering terms:
\[
\text{eusup}_{ij} = \alpha_0 + \ldots + \alpha_1 \text{GSPEND}_j + \gamma_0 lft_{ij}
\]
\[\quad + \gamma_1 lft_{ij} \times \text{GSPEND}_j + \ldots + \left( u_j^0 + lft_{ij} u_j^1 + \epsilon_{ij} \right)
\]
\[\Rightarrow \quad \frac{\partial \text{eusup}}{\partial lft_{ij}} = b_{lft_{ij}} + b_{lGS} \text{GSPEND} + u_j^1 \quad \& \quad \frac{\partial \text{eusup}}{\partial \text{GSPEND}} = b_{GS} + b_{lGS} \text{edu}
\]
\[E \left( \frac{\partial \text{eusup}}{\partial lft_{ij}} \right) = b_{lft_{ij}} + b_{lGS} \text{GSPEND}
\]
Properties of OLS under HLM Conditions

- Properties of OLS Estimates of Lin-Interact Model if truly RE/HLM:
  \[ y = \beta_0 + \beta_x x + \beta_z z + \beta_{xz} xz + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = X\beta + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) \]

- So, OLS coeff. est’s sill differ from truth by \( A\varepsilon^* \):
  \[ \hat{\beta}_{\text{LS}} = (X'X)^{-1} X'y = (X'X)^{-1} X' \left[ X\beta + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) \right] \]
  \[ = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = \beta + (X'X)^{-1} X'\varepsilon^* \]

- So, OLS coeff. est’s unbiased & consistent:
  \[ E(\hat{\beta}_{\text{LS}}) = E[\beta + (X'X)^{-1} X'\varepsilon^*] = E[\beta + (X'X)^{-1} X'(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)] \]
  \[ = \beta + (X'X)^{-1} X'E(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = \beta + (X'X)^{-1} X'[E(\varepsilon^0) + E(\varepsilon^1)x + E(\varepsilon^2)z] \]
  \[ = \beta + (X'X)^{-1} X'[0 + E(\varepsilon^1)x + E(\varepsilon^2)z] = \beta + (X'X)^{-1} X'[0 + 0 + 0] = \beta. \quad Q.E.D. \]

- Note: only works for models w/ additively separable stochastic component; not nec’ly for others (log/prob)
Properties of OLS under HLM Conditions

- But, OLS s.e.’s will be wrong; not \( s^2(X'X)^{-1} \), but:

\[
V(\hat{\beta}_{LS}) = V\left[ \beta + (X'X)^{-1} X'\varepsilon^* \right] = V[\beta] + V\left[ (X'X)^{-1} X'\varepsilon^* \right] + 2C\left[ \beta, (X'X)^{-1} X'\varepsilon^* \right] = 0 + V\left[ (X'X)^{-1} X'\varepsilon^* \right] + 0 = (X'X)^{-1} X'V(\varepsilon^*)X(X'X)^{-1} = (X'X)^{-1} X'\left[ V(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) \right]X(X'X)^{-1} = (X'X)^{-1} X'\left[ V(\varepsilon^0) + V(\varepsilon^1 x) + V(\varepsilon^2 z) \right]X(X'X)^{-1}
\]

(the covariance terms are assumed zero)
Sandwich Estimators

\[ V(\hat{\beta}_{\text{LS}}) = (X'X)^{-1}X'[V(\varepsilon^0) + V(\varepsilon'x) + V(\varepsilon^2z)]X(X'X)^{-1} \]

- Not \( \sigma^2 I \) (even if each \( \varepsilon^* \) is), so whole thing doesn’t reduce to \( \sigma^2(X'X)^{-1} \), so OLS s.e.’s wrong.
- Be OK on avg (unbiased) & in limit (consistent) if that term varied in way “orthogonal to \( XX' \)”
  - But def’ly not b/c \( \cdot \) includes \( x \) & \( z \), which part of \( X! \)
  - =brilliant insight of ‘robust’ (i.e., consistent) s.e. est’s:
    - Only need s.e. formula that accounts relation \( V(\varepsilon^*) \) to “\( X'X \)”, i.e., regressors, squares, & cross-prod’s involved in \( X'[\cdot]X' \)
- \( \Rightarrow " \), robust” & “, cluster” can work (for RE & HLM, resp’ly)
  - \( \hat{V}(\beta)_{\text{RE}} = \sigma^2(I + xx' + zz') \) so track \( e^2 \) rel \( xx' \) & \( zz' \) \( \Rightarrow \)
  - \( \hat{V}(\beta)_{\text{HM}} = \sigma_0^2 I + \sigma_1^2 xx' + \sigma_2^2 zz' \) sim + grpgng \( \Rightarrow \)
  \[
  [\cdot] = \frac{1}{n} \sum_{i=1}^{n} e_i^2 X_i X_i' \]
  \[
  [\cdot] = \sum_{j=1}^{J} \left\{ \frac{n_j}{(\sum_{i=1}^{n_j} e_{ij} X_{ij})'} (\sum_{i=1}^{n_j} e_{ij} X_{ij}) \right\}
  \]
From the CLRM to HLM

\[
y = \beta_0 + \alpha_0 x + \gamma_0 z + (\alpha_1 + \gamma_1) xz + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)
\]

\[
V(\hat{\beta}_{LS}) = (X'X)^{-1} X' [V(\varepsilon^0) + V(\varepsilon^1 x) + V(\varepsilon^2 z)] X (X'X)^{-1}
\]

- ⇒ appropriate “, robust” & “, cluster” can work
  - I.e., asymptotically std errs right...BUT need large \(n_j\)
  - I.e., coefficients still inefficient.
    - Want/need efficiency, or \(n_j\) low? HLM/RE or FGLS/FWLS.
    - Note: similarity RE/RC and HLM, RE & FWLS. As suggests, RE only helps efficiency and only rightly does so if that’s all it does. (I.e., if the RE’s orthogonal to X.)
  - I.e., “work” thusly for models with additively-separable stochastic components
    - As w/ all such “sandwich” estimators, logical disconnect in applying them to models w/o such separability.
Properties of OLS under HLM Conditions

Properties of OLS Estimates of Lin-Interact Model if truly RE/HLM:

\[
y = \beta_0 + \beta_x x + \beta_z z + \beta_{xz} xz + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = X\beta + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)
\]

So, OLS coeff. est’s sill differ from truth by \(A\varepsilon^*\):

\[
\hat{\beta}_{LS} = (X'X)^{-1} X' y = (X'X)^{-1} X' [X\beta + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)] \\
= (X'X)^{-1} X' X\beta + (X'X)^{-1} X' (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = \beta + (X'X)^{-1} X' \varepsilon^*
\]

So, OLS coeff. est’s unbiased & consistent (iff...):

\[
E(\hat{\beta}_{LS}) = E[\beta + (X'X)^{-1} X' \varepsilon^*] = E[\beta + (X'X)^{-1} X' (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z)] \\
= \beta + (X'X)^{-1} X' E(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) = \beta + (X'X)^{-1} X' [E(\varepsilon^0) + E(\varepsilon^1) x + E(\varepsilon^2) z] \\
= \beta + (X'X)^{-1} X' [0 + E(\varepsilon^1) x + E(\varepsilon^2) z] = \beta + (X'X)^{-1} X' [0 + 0 + 0] = \beta. \quad Q.E.D.
\]

Note: only works for models with additively separable stochastic component; not nec’ly others (e.g., logit/probit)
Properties of OLS under HLM Conditions

- But, OLS s.e.’s will be wrong; not \( s^2(X'X)^{-1} \), but:

\[
V(\hat{\beta}_{LS}) = V\left[ \beta + (X'X)^{-1} X'\epsilon^* \right]
\]

\[
= V[\beta] + V\left[ (X'X)^{-1} X'\epsilon^* \right] + 2C\left[ \beta, (X'X)^{-1} X'\epsilon^* \right]
\]

\[
= 0 + V\left[ (X'X)^{-1} X'\epsilon^* \right] + 0
\]

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

\[
= (X'X)^{-1} X'\left[ V(\epsilon^0 + \epsilon^1 x + \epsilon^2 z) \right]X(X'X)^{-1}
\]

\[
= (X'X)^{-1} X'\left[ V(\epsilon^0) + V(\epsilon^1 x) + V(\epsilon^2 z) \right]X(X'X)^{-1}
\]

*(note: the covariance terms are assumed zero)*
Sandwich Estimators

\[ V(\hat{\beta}_{LS}) = (X'X)^{-1}X'[V(\varepsilon^0) + V(\varepsilon'x) + V(\varepsilon^2z)]X(X'X)^{-1} \]

- Not \( \sigma^2I \) (even if each \( \varepsilon^* \) is), so whole thing doesn’t reduce to \( \sigma^2(X'X)^{-1} \), so OLS s.e.’s wrong.
- Be OK on avg (unbiased) & in limit (consistent) if that term varied in way “orthogonal to \( xx' \)”
  - But def’ly not b/c \([ \cdot ]\) includes \( x \) & \( z \), which part of \( X \)!
  - =brilliant insight of ‘robust’ (i.e., consistent) s.e. est’s:
    - Only need s.e. formula that accounts relation \( V(\varepsilon^*) \) to “\( X'X \)”, i.e., regressors, squares, & cross-prod’s involved in \( X'[\cdot]X \)”
- \( \Rightarrow “, robust” \& “, cluster” can work (for RE & HLM, resp’ly)
- \( \hat{V}(\hat{\beta})_{RE} = \sigma^2(I + xx' + zz') \) so track \( e^2 \) rel \( xx' \) & \( zz' \) \( \Rightarrow \) heterosked.: \\
  \[ [. \cdot] = \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i x_i' \]
- \( \hat{V}(\hat{\beta})_{HM} = \sigma_0^2I + \sigma_1^2 xx' + \sigma_2^2 zz' \) sim+grpng \( \Rightarrow \) cluster: \\
  \[ [. \cdot] = \sum_{j=1}^{J} \left\{ \frac{n_j}{\sum_{i=1}^{n_j} e_{ij} X_{ij}} (\sum_{i=1}^{n_j} e_{ij} X_{ij}) \right\} \]
From the CLRM to HLM

\[ y = \beta_0 + \alpha_0 x + \gamma_0 z + (\alpha_1 + \gamma_1) x z + (\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z) \]

\[ V(\hat{\beta}_{LS}) = (X'X)^{-1} X' \left[ V(\varepsilon^0) + V(\varepsilon^1 x) + V(\varepsilon^2 z) \right] X(X'X)^{-1} \]

- \Rightarrow\text{ appropriate “, robust” & “, cluster” can work}
  - I.e., \textit{asymptotically} s.e.’s right...BUT need large \(n_j\) & \(N-k\)
  - Specifically, from small-sample corrections, seems need small:
    \[
    \left[ \frac{n_j}{(n_j - 1)} \right] \left[ \frac{(N - 1)}{(N - k)} \right]
    \]
  - I.e., \textit{coefficients still inefficient}.
    - Want/need efficiency, or \(n_j\) or \(N-k\) low? HLM/RE or FGLS/FWLS.
    - Note: similarity RE/RC and HLM, RE & FWLS. As that sim suggests, RE/RC only helps efficiency and only rightly does so if that’s all it does. (I.e., if the RE/RC’s orthogonal to \(X\).)
  - I.e., “work” thusly only or models with additively-separable stochastic components
    - As w/ all such “sandwich” estimators, a sort of logical disconnect in applying them to models w/o such separability...