From CLRM to Multilevel Model

$$\text{eusup}_{ij} = \beta_j^0 + \beta_j^{lr}\text{lftrt}_{ij} + \ldots + \varepsilon_{ij}$$

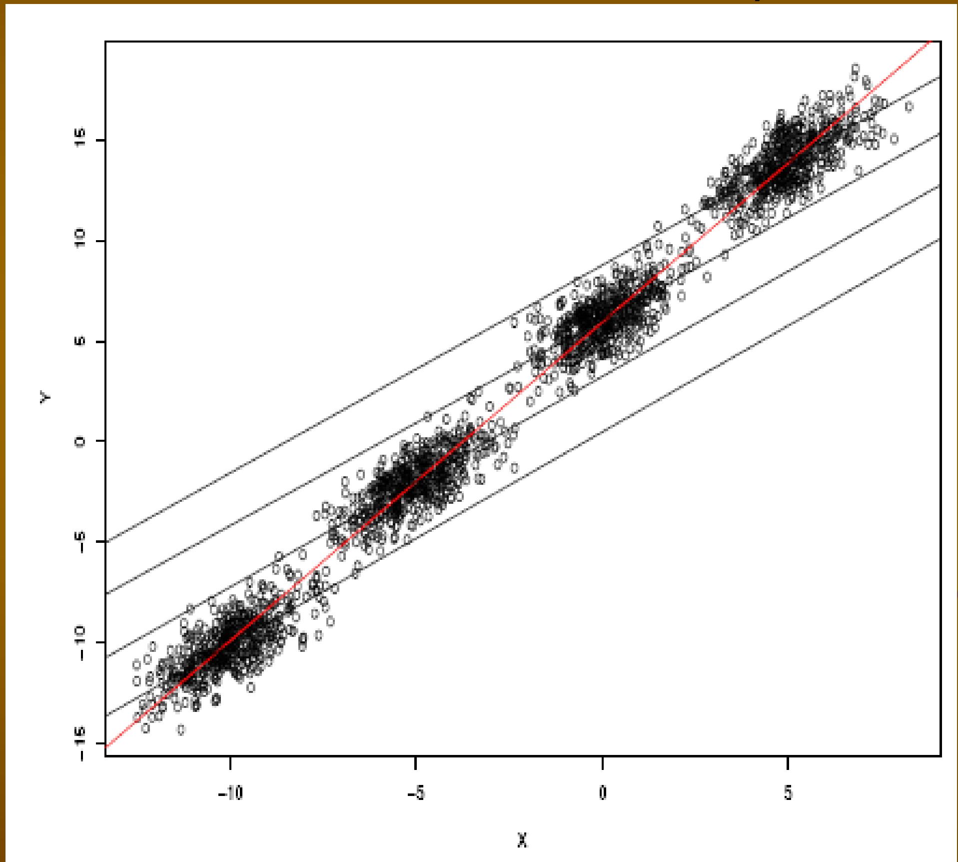
$$\beta_j^0 = \alpha_0 + \alpha_1 GSPEND_j + u_j^0$$

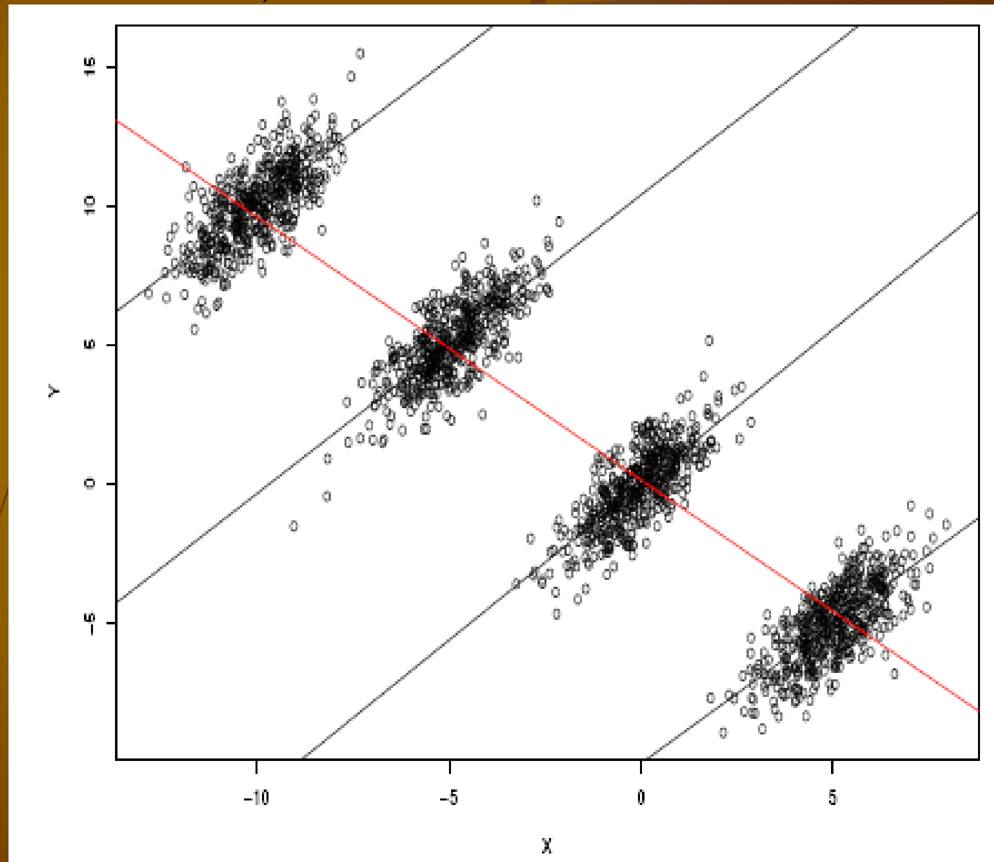
$$\beta_j^{lr} = \gamma_0 + \gamma_1 GSPEND_j + u_j^1$$

- 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 hetereogeneity 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)+error*: $y_i = \beta_0 + \beta_x x_i + \beta_z z_i + \beta_{xz} x_i z_i + \varepsilon_i$
  - But *dy/dx=f(z)* w/o error!: $\frac{\partial y}{\partial x} = \beta_x + \beta_{xz} xz$
  - So, try:

$$y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon^0$$
$$\frac{\partial y}{\partial x} \equiv \beta_1 = \alpha_0 + \alpha_1 z + \varepsilon^1$$
$$\frac{\partial y}{\partial z} \equiv \beta_2 = \gamma_0 + \gamma_1 x + \varepsilon^2$$
$$\Rightarrow y = \beta_0 + \left(\alpha_0 + \alpha_1 z + \varepsilon^1\right) x + \left(\gamma_0 + \gamma_1 x + \varepsilon^2\right) z + \varepsilon^0$$
$$= \beta_0 + \alpha_0 x + \gamma_0 z + \left(\alpha_1 + \gamma_1\right) xz + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)$$

  - => 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.

$$\varepsilon^* = \varepsilon_{ij}^0 + \varepsilon_j^1 x_{ij} + \varepsilon_{ij}^2 z_j$$

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.

$$\text{eusup}_{ij} = \beta_j^0 + \beta_j^{lr}\text{lftrt}_{ij} + \ldots + \varepsilon_{ij}$$

$$\beta_j^0 = \alpha_0 + \alpha_1 GSPEND_j + u_j^0$$

$$\beta_j^{lr} = \gamma_0 + \gamma_1 GSPEND_j + u_j^1$$

$$\Rightarrow \text{eusup}_{ij} = \alpha_0 + \alpha_1 GSPEND_j + u_j^0 + \gamma_0 \text{lftrt}_{ij}$$

$$+\gamma_1 \text{lftrt}_{ij} \times GSPEND_j + \text{lftrt}_{ij} u_j^1 + \ldots + \varepsilon_{ij}$$

gathering terms:

$$\text{eusup}_{ij} = \alpha_0 + \ldots + \alpha_1 GSPEND_j + \gamma_0 \text{lftrt}_{ij}$$

$$+\gamma_1 \text{lftrt}_{ij} \times GSPEND_j + \ldots + \left(u_j^0 + \text{lftrt}_{ij} u_j^1 + \varepsilon_{ij}\right)$$

$$\Rightarrow \frac{\partial \text{eusup}}{\partial \text{lftrt}} = b_{lftrt} + b_{lrGS}\text{GS} + u_j^1 \quad \& \quad \frac{\partial \text{eusup}}{\partial \text{GS}} = b_{GS} + b_{lrGS}\text{edu}$$

$$E\left(\frac{\partial \text{eusup}}{\partial \text{lftrt}}\right) = b_{lftrt} + b_{lrGS}\text{GS}$$

Properties of OLS under HLM Conditions

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

$$y = \beta_0 + \beta_x x + \beta_z z + \beta_{xz} xz + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \mathbf{X\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)$$

- So, OLS coeff. est's sill differ from truth by $\mathbf{A}\varepsilon^*$:

$$\begin{aligned}\hat{\boldsymbol\beta}_{\mathbf{LS}} &= (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}\left[\mathbf{X\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= (\mathbf{X'X})^{-1}\mathbf{X'X\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon^*\end{aligned}$$

- So, OLS coeff. est's unbiased & consistent:

$$\begin{aligned}E\left(\hat{\boldsymbol\beta}_{\mathbf{LS}}\right) &= E\left[\boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\varepsilon^*\right] = E\left[\boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= \boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}E\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\left[E\left(\varepsilon^0\right) + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] \\ &= \boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}\left[0 + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] = \boldsymbol\beta + (\mathbf{X'X})^{-1}\mathbf{X'}[0+0+0] = \boldsymbol\beta. \quad Q.E.D.\end{aligned}$$

  - 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:

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

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

$$=0+V\left[(\mathbf{X'X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}^*\right]+0$$

$$=(\mathbf{X'X})^{-1}\mathbf{X}'V\left(\boldsymbol{\varepsilon}^*\right)\mathbf{X}(\mathbf{X'X})^{-1}$$

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

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

(the covariance terms are assumed zero)

Sandwich Estimators

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

- Not $\sigma^2 I$ (even if each $\boldsymbol\varepsilon^*$ 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 [·] 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($\boldsymbol\varepsilon^*$) to "X'X", i.e., regressors, squares, & cross-prod's involved in X'[·]X"
- ⇒ ", robust" & ", cluster" can work (for RE & HLM, resp'ly)
  - $\hat{V}\left(\hat{\boldsymbol\beta}\right)_{RE} = \sigma^2\left(\mathbf{I} + \boldsymbol{xx}' + \boldsymbol{zz}'\right)$ so track e² rel XX' & ZZ' ⇒ $[\cdot] = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{X_i X_i}'$
  - $\hat{V}\left(\hat{\boldsymbol\beta}\right)_{HM} = \sigma_0^2\mathbf{I} + \sigma_1^2\boldsymbol{xx}' + \sigma_2^2\boldsymbol{zz}'$ sim + grpng⇒ h $[\cdot] = \sum_{j=1}^{J}\left\{(\sum_{i=1}^{n_j} e_{ij}\boldsymbol{X_{ij}})'(\sum_{i=1}^{n_j} e_{ij}\boldsymbol{X_{ij}})\right\}$

From the CLRM to HLM

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

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

- ⇒ appropriate ", robust" & ", cluster" can work
  - I.e., <u>*asymptotically*</u> std errs right…BUT need large n_j
  - I.e., <u>*coefficients still inefficient*</u>.
    - 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:

$$y = \beta_0 + \beta_x x + \beta_z z + \beta_{xz} xz + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \mathbf{X}\boldsymbol{\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)$$

- So, OLS coeff. est's sill differ from truth by $\mathbf{A}\boldsymbol{\varepsilon}^*$:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathbf{LS}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[\mathbf{X}\boldsymbol{\beta} + \left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon^*\end{aligned}$$

- So, OLS coeff. est's unbiased & consistent (*iff*...):

$$\begin{aligned}E\left(\hat{\boldsymbol{\beta}}_{\mathbf{LS}}\right) &= E\left[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon^*\right] = E\left[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right)\right] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left(\varepsilon^0 + \varepsilon^1 x + \varepsilon^2 z\right) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[E\left(\varepsilon^0\right) + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[0 + E\left(\varepsilon^1\right)x + E\left(\varepsilon^2\right)z\right] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[0+0+0] = \boldsymbol{\beta}. \quad \textit{Q.E.D.}\end{aligned}$$

  - 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 $s^2(\mathbf{X}'\mathbf{X})^{-1}$, but:

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

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

$$=0+V\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}^{*}\right]+0$$

$$=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'V\left(\boldsymbol{\varepsilon}^{*}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

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

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

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

Sandwich Estimators

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

- Not $\sigma^2 I$ (even if each $\boldsymbol{\varepsilon}^*$ is), so whole thing doesn't reduce to $\sigma^2(\mathbf{X'X})^{-1}$, so OLS s.e.'s wrong.
- Be OK on avg (unbiased) & in limit (consistent) if that term varied in way " orthogonal to $\mathbf{xx'}$ "
  - But def'ly not b/c [·] 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($\boldsymbol{\varepsilon}^*$) to "**X'X**", i.e., regressors, squares, & cross-prod's involved in **X'**[·]**X**"
- ⇒ ", robust" & ", cluster" can work (for RE & HLM, resp'ly)
  - $\hat{V}\left(\hat{\boldsymbol{\beta}}\right)_{RE}=\sigma^2\left(\mathbf{I}+\boldsymbol{xx}'+\boldsymbol{zz}'\right)$ so track e^2 rel $\boldsymbol{xx}'$ & $\mathbf{zz'}$ ⇒ *heterosked.*: $[\cdot]=\frac{1}{n}\sum_{i=1}^{n}e_i^2\boldsymbol{X}_i\boldsymbol{X}_i'$
  - $\hat{V}\left(\hat{\boldsymbol{\beta}}\right)_{HM}=\sigma_0^2\mathbf{I}+\sigma_1^2\boldsymbol{xx}'+\sigma_2^2\boldsymbol{zz}'$ sim+grpng⇒ *cluster*: $[\cdot]=\sum_{j=1}^{J}\left\{(\sum_{i=1}^{n_j}e_{ij}\boldsymbol{X}_{ij})'(\sum_{i=1}^{n_j}e_{ij}\boldsymbol{X}_{ij})\right\}$

From the CLRM to HLM

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

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

- $\Rightarrow$ appropriate ", robust" & ", cluster" can work
  - I.e., ***asymptotically*** s.e.'s right…BUT need large n_j & N-k
    - Specifically, from small-sample corrections, seems need small: $\left[n_j/(n_j-1)\right]\left[(N-1)/(N-k)\right]$
  - I.e., ***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…