Generalized Marginal Models for Longitudinal Data (Inference and Analysis)

Recall. . .

We specified a **generalized linear marginal model (GLMM)** by using a **link function** which connects the **conditional mean** to a **linear predictor**, alongside a **variance function** which depends on the mean, and **pairwise association** matrix.

That is, $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$ with $\text{var}(Y_{ij}) = \phi V(\mu_{ij})$ and $\text{cor}(Y_{ij}, Y_{i\ell}) = \mathbf{R}_i(\rho)$.

How do we estimate these models?

Further recall. . .

That if $U(\theta, \mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n} \Psi_i(Y_i, \theta)$, with $U(\widehat{\theta}, \mathbf{Y}) = 0$, then $\widehat{\theta}$ is consistent (and asymptotically normal) for θ_0 where θ_0 is such that $E[U(\theta_0, \mathbf{Y})] = 0$.

This is an **M-estimator** and it carries with it nice properties!

What if **we** formed an **M-estimator** based on our assumptions from the **GLMM**?

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.
- This is equivalently, $\mu_{ij} = E[Y_{ij}|X_{ij}] = g^{-1}(X_{ij}\beta)$.

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.
- This is equivalently, $\mu_{ij} = E[Y_{ij}|X_{ij}] = g^{-1}(X_{ij}\beta)$.
- Define $\Psi_{ij}(Y_{ij}, X_{ij}, \beta) = Y_{ij} - g^{-1}(X_{ij}\beta)$.

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.
- This is equivalently, $\mu_{ij} = E[Y_{ij}|X_{ij}] = g^{-1}(X_{ij}\beta)$.
- Define $\Psi_{ij}(Y_{ij}, X_{ij}, \beta) = Y_{ij} - g^{-1}(X_{ij}\beta)$.
- This is such that $E[\Psi_{ij}(Y_{ij}, X_{ij}, \beta)] = 0$ at the **true** β.

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.
- This is equivalently, $\mu_{ij} = E[Y_{ij}|X_{ij}] = g^{-1}(X_{ij}\beta)$.
- Define $\Psi_{ij}(Y_{ij}, X_{ij}, \beta) = Y_{ij} - g^{-1}(X_{ij}\beta)$.
- This is such that $E[\Psi_{ij}(Y_{ij}, X_{ij}, \beta)] = 0$ at the **true** β.
- This suggests using

$$\sum_{i=1}^{n} \left\{ Y_{ij} - g^{-1}(X_{ij}\beta) \right\}.$$

How might we do this?

- We have specified that $g(E[Y_{ij}|X_{ij}]) = X_{ij}\beta$.
- This is equivalently, $\mu_{ij} = E[Y_{ij}|X_{ij}] = g^{-1}(X_{ij}\beta)$.
- Define $\Psi_{ij}(Y_{ij}, X_{ij}, \beta) = Y_{ij} - g^{-1}(X_{ij}\beta)$.
- This is such that $E[\Psi_{ij}(Y_{ij}, X_{ij}, \beta)] = 0$ at the **true** β.
- This suggests using
$$\sum_{i=1}^{n} \left\{ Y_{ij} - g^{-1}(X_{ij}\beta) \right\}.$$
- We could equivalently do this as
$$\sum_{i=1}^{n} \left\{ Y_{i} - g^{-1}(X_{i}\beta) \right\}.$$

Generalized Estimating Equations (GEE)

- This approach *would* work!

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1} (Y_i - g^{-1}(X_i \beta)).$$

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - g^{-1}(X_i\beta)).$$

- Here,

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - g^{-1}(X_i\beta)).$$

- Here,
  - $D_i = \frac{\partial}{\partial \beta} g^{-1}(X_i\beta)$ (a $k \times p$ matrix).

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - g^{-1}(X_i\beta)).$$

- Here,
  - $D_i = \frac{\partial}{\partial\beta} g^{-1}(X_i\beta)$ (a $k \times p$ matrix).
  - $V_i = \phi A_i^{1/2} \mathbf{R}_i(\rho) A_i^{1/2}$ is the **working covariance matrix** ($k \times k$).

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - g^{-1}(X_i\beta)).$$

- Here,
  - $D_i = \frac{\partial}{\partial\beta} g^{-1}(X_i\beta)$ (a $k \times p$ matrix).
  - $V_i = \phi A_i^{1/2} \mathbf{R}_i(\rho) A_i^{1/2}$ is the **working covariance matrix** ($k \times k$).
  - V_i will generally be a function of $\mu_i = g^{-1}(X_i\beta)$.

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - g^{-1}(X_i\beta)).$$

- Here,
  - $D_i = \frac{\partial}{\partial\beta} g^{-1}(X_i\beta)$ (a $k \times p$ matrix).
  - $V_i = \phi A_i^{1/2} \mathbf{R}_i(\rho) A_i^{1/2}$ is the **working covariance matrix** ($k \times k$).
  - V_i will generally be a function of $\mu_i = g^{-1}(X_i\beta)$.
- Note, this is *essentially* just QMLE!

Generalized Estimating Equations (GEE)

- This approach *would* work!
- This is not going to be the **most efficient** estimator.
- We turn to **Generalized Estimating Equations** (GEE), which define

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1} (Y_i - g^{-1}(X_i \beta)).$$

- Here,
  - $D_i = \frac{\partial}{\partial \beta} g^{-1}(X_i \beta)$ (a $k \times p$ matrix).
  - $V_i = \phi A_i^{1/2} \mathbf{R}_i(\rho) A_i^{1/2}$ is the **working covariance matrix** ($k \times k$).
  - V_i will generally be a function of $\mu_i = g^{-1}(X_i \beta)$.
- Note, this is *essentially* just QMLE!
- We will see some specific examples in the next lecture!

Asymptotic Inference of GEE

From this estimator we get $\widehat{\beta} \dot{\sim} N(\beta, \text{var}(\widehat{\beta}))$. $\text{var}(\widehat{\beta}) = J^{-1}\Gamma J^{-1\prime}$, as we saw for **M-estimators** generally.

$$J = E\left[-\frac{\partial}{\partial\beta}U(\beta)\right] = \sum_{i=1}^{n} D_i' V_i^{-1} D_i$$

$$\Gamma = E\left[U(\beta)U(\beta)'\right] = \sum_{i=1}^{n} D_i' V_i^{-1} \text{var}(Y_i) V_i^{-1} D_i$$

Scale Parameters

Typically ρ and ϕ are considered **nuisance** parameters.

In the **GEE method** they are estimated **ad-hoc** based on formulas from **Pearson Residuals**.

The *specific* formula for $\widehat{\rho}$ and $\widehat{\phi}$ are **unimportant**, typically.

$$\widehat{r}_{ij} = \frac{Y_{ij} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu_{ij}})}}$$

$$\widehat{\phi} = \frac{1}{N-p} \sum_{i=1}^{n} \sum_{j=1}^{k_i} \widehat{r}_{ij}^2$$

$$\widehat{\rho}_{jk} = \frac{1}{\widehat{\phi}(n-p)} \sum_{i=1}^{n} \widehat{r}_{ij} \widehat{r}_{ik}.$$

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.
   - $U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1} C\mathbf{G}_m U(\widetilde{\beta}_G) \sim \chi^2_r$

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.
   - $U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1} C\mathbf{G}_m U(\widetilde{\beta}_G) \sim \chi^2_r$
   - $\widetilde{\beta}_G$ is the GEE estimator assuming H_0.

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.
   - $U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1} C\mathbf{G}_m U(\widetilde{\beta}_G) \sim \chi^2_r$
   - $\widetilde{\beta}_G$ is the GEE estimator assuming H_0.
   - $\mathbf{G}_m = \widehat{J}^{-1}$ and $\mathbf{G}_r = \widehat{J}^{-1}\widehat{\Gamma}\widehat{J}^{-1}$.

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.
   - $U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1} C\mathbf{G}_m U(\widetilde{\beta}_G) \sim \chi^2_r$
   - $\widetilde{\beta}_G$ is the GEE estimator assuming H_0.
   - $\mathbf{G}_m = \widehat{J}^{-1}$ and $\mathbf{G}_r = \widehat{J}^{-1}\widehat{\Gamma}\widehat{J}^{-1}$.
   - You do not need to learn this!

Confidence Intervals and Hypothesis Tests

1. Wald procedures based on asymptotic normality.
2. **Generalized score statistics**, which extends the LRT to GEEs.
   - Test $H_0 : C\beta = 0$.
   - $U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1} C\mathbf{G}_m U(\widetilde{\beta}_G) \sim \chi^2_r$
   - $\widetilde{\beta}_G$ is the GEE estimator assuming H_0.
   - $\mathbf{G}_m = \widehat{J}^{-1}$ and $\mathbf{G}_r = \widehat{J}^{-1}\widehat{\Gamma}\widehat{J}^{-1}$.
   - You do not need to learn this!
3. Can also modify `AIC` to produce `QIC`.

Inference is **valid** even if V_i is incorrectly specified.

Why do we bother?

Parameter Interpretation

Parameters in **marginal models** are interpreted as **population-level effects**.

That is, β_j is the impact of variate j *on average*, across the whole population (assuming other variates fixed).

These are **not** individual level effects!

A Note on Time-Varying Covariates

The linear structure of our marginal models **implicitly assumes** that $Y_{ij} \perp X_{ik} | X_{ij}$, for all $j \neq k$.

That is, the **outcome** is *conditionally* independent of explanatory factors at **all other times**, given the **current time's** explanatory factors.

A Note on Time-Varying Covariates

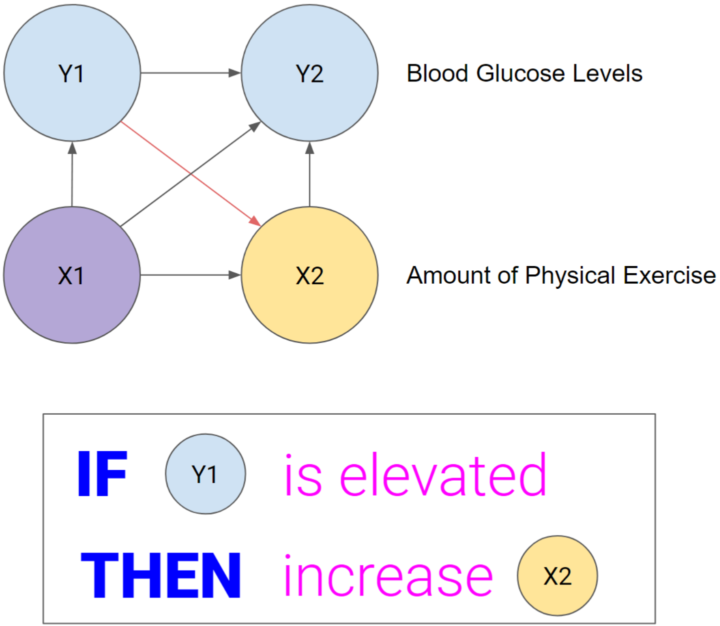
- This will generally be true of **time invariant** variables.

A Note on Time-Varying Covariates

- This will generally be true of **time invariant** variables.
- This will also generally be true of **deterministic, time-varying** variables.

A Note on Time-Varying Covariates

- This will generally be true of **time invariant** variables.
- This will also generally be true of **deterministic, time-varying** variables.
- This will not be generally true of **stochastic, time-varying** variables.


Y1
Y2
Blood Glucose Levels
X1
X2
Amount of Physical Exercise
IF
Y1
is elevated
THEN
increase
X2

Summary

- We can use **generalized estimating equations** to solve for the β parameters in a marginal model.

Summary

- We can use **generalized estimating equations** to solve for the β parameters in a marginal model.
- This procedures results in **asymptotically normal** estimators, allowing for standard inference.

Summary

- We can use **generalized estimating equations** to solve for the β parameters in a marginal model.
- This procedures results in **asymptotically normal** estimators, allowing for standard inference.
- The parameters are **population averaged** effects, *not* individual.

Summary

- We can use **generalized estimating equations** to solve for the β parameters in a marginal model.
- This procedures results in **asymptotically normal** estimators, allowing for standard inference.
- The parameters are **population averaged** effects, *not* individual.
- We need to be mindful of **stochastic, time-varying** covariates.