Qian, Klasnja, and Murphy's Linear Mixed Effects Model with Endogenous Covariates

Hunyong Cho, Josh Zitovsky, Matthew Brown, Xinyi Li, Minxin Lu, Kushal Shah, John Sperger, and Michael Kosorok

University of North Carolina at Chapel Hill

1/24/2020

Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study

Tianchen Qian, Predrag Klasnja and Susan A. Murphy

Department of Statistics, Harvard University, Cambridge, MA 02138
e-mail: qiantianchen@fas.harvard.edu; samurphy@fas.harvard.edu.

School of Information, University of Michigan, Ann Arbor, MI 48109
e-mail: klasnja@umich.edu.

Abstract: Mobile health is a rapidly developing field in which behavioral treatments are delivered to individuals via wearables or smartphones to facilitate health-related behavior change. Micro-randomized trials (MRT) are an experimental design for developing mobile health interventions. In an MRT the treatments are randomized numerous times for each individual over course of the trial. Along with assessing treatment effects, behavioral scientists aim to understand between-person heterogeneity in the treatment effect. A natural approach is the familiar linear mixed model. However, directly applying linear mixed models is problematic because potential moderators of the treatment effect are frequently endogenous—that is, may depend on prior treatment. We discuss model interpretation and biases that arise in the absence of additional assumptions when endogenous covariates are included in a linear mixed model. In particular, when there are endogenous covariates, the coefficients no longer have the customary marginal interpretation. However, these coefficients still have a conditional-on-the-random-effect interpretation. We provide an additional assumption that, if true, allows scientists to use standard software to fit linear mixed model with endogenous covariates, and person-specific predictions of effects can be provided. As an illustration, we assess the effect of activity suggestion in the HeartSteps MRT and analyze the between-person treatment effect heterogeneity.

- 1. Classic Linear Mixed Effects Models (LMM)
- 2. QKM's LMM for endogenous covariate data
- 3. Our discussion

1. Classic Linear Mixed Effects Models

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n;\ t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n; \; t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

One more important assumption:

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n;\ t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

One more important assumption:

- $X_{i,t}$ is {fixed},

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n;\ t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

One more important assumption:

- $X_{i,t}$ is {fixed},
- or {exogenous & independent of b_i}.

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n;\ t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

One more important assumption:

- $X_{i,t}$ is $\{$fixed$\}$,
- or $\{$exogenous & independent of $b_i\}$.

Thanks to the independence ($X_{i,t} \perp b_i$), β has a marginal interpretation!

Classic Linear Mixed Effects Models

Classic linear mixed effects models (LMM)

- $Y_{i,t+1} = X_{i,t}\beta + Z_{i,t}b_i + \epsilon_{i,t+1}, \qquad i = 1, 2, .., n; \; t = 1, 2, ..., T$
- $\epsilon_{i,t+1} \sim N(0, \sigma^2 I)$
- $b_i \sim N(0, G)$

One more important assumption:

- $X_{i,t}$ is {fixed},
- or {exogenous & independent of b_i}.

Thanks to the independence ($X_{i,t} \perp b_i$), β has a marginal interpretation!

- from $E[Y_{i,t}|X_{i,t}, b_i] = X_{i,t}\beta + Z_{i,t}b_i$, we get
$$E[Y_{i,t}|X_{i,t}] = X_{i,t}\beta + \underbrace{E[Z_{i,t}b_i]}_{=Z_{i,t}E[b_i|X_i]=Z_{i,t}E[b_i]=0} = X_{i,t}\beta.$$

Challenge when using LMM in Precision Medicine

Consider a micro-randomized trial (MRT).

- We need treatment variable ($A_{i,t}$):

$$Y_{i,t+1} = X_{i,t}\beta_0 + A_{i,t}X_{i,t}\beta_1 + Z_{i,t}b_{0,i} + A_{i,t}Z_{i,t}b_{1,i} + \epsilon_{i,t+1}.$$

Challenge when using LMM in Precision Medicine

Consider a micro-randomized trial (MRT).

- We need treatment variable ($A_{i,t}$):

$$Y_{i,t+1} = X_{i,t}\beta_0 + A_{i,t}X_{i,t}\beta_1 + Z_{i,t}b_{0,i} + A_{i,t}Z_{i,t}b_{1,i} + \epsilon_{i,t+1}.$$

- Want to use historical information ($H_{i,t}$) for the next treatment, not only the current covariate ($X_{i,t}$).

$$H_{i,t} = (X_{i,1}, A_{i,1}, Y_{i,2}, ..., X_{i,t-1}, A_{i,t-1}, Y_{i,t}, X_{i,t}).$$

Challenge when using LMM in Precision Medicine

Consider a micro-randomized trial (MRT).

- We need treatment variable ($A_{i,t}$):

$$Y_{i,t+1} = X_{i,t}\beta_0 + A_{i,t}X_{i,t}\beta_1 + Z_{i,t}b_{0,i} + A_{i,t}Z_{i,t}b_{1,i} + \epsilon_{i,t+1}.$$

- Want to use historical information ($H_{i,t}$) for the next treatment, not only the current covariate ($X_{i,t}$).

$$H_{i,t} = (X_{i,1}, A_{i,1}, Y_{i,2}, ..., X_{i,t-1}, A_{i,t-1}, Y_{i,t}, X_{i,t}).$$

- $Y_{i,t+1} = f_0(H_{i,t})^\top \beta_0 + A_{i,t} f_1(H_{i,t})^\top \beta_1 + g_0(H_{i,t})^\top b_{0,i} + A_{i,t} g_1(H_{i,t})^\top b_{1,i} + \epsilon_{i,t+1}.$
  $f.$ and $g.$ are some summary functions with a fixed dimension.

Challenge when using LMM in Precision Medicine

Consider a micro-randomized trial (MRT).

- We need treatment variable ($A_{i,t}$):

$$Y_{i,t+1} = X_{i,t}\beta_0 + A_{i,t}X_{i,t}\beta_1 + Z_{i,t}b_{0,i} + A_{i,t}Z_{i,t}b_{1,i} + \epsilon_{i,t+1}.$$

- Want to use historical information ($H_{i,t}$) for the next treatment, not only the current covariate ($X_{i,t}$).

$$H_{i,t} = (X_{i,1}, A_{i,1}, Y_{i,2}, ..., X_{i,t-1}, A_{i,t-1}, Y_{i,t}, X_{i,t}).$$

- $Y_{i,t+1} = f_0(H_{i,t})^\top \beta_0 + A_{i,t} f_1(H_{i,t})^\top \beta_1 + g_0(H_{i,t})^\top b_{0,i} + A_{i,t} g_1(H_{i,t})^\top b_{1,i} + \epsilon_{i,t+1}$.
  $f.$ and $g.$ are some summary functions with a fixed dimension.
- Is there any problem using LMM for this model?

Challenge when using LMM in Precision Medicine

Consider a micro-randomized trial (MRT).

- We need treatment variable ($A_{i,t}$):

$$Y_{i,t+1} = X_{i,t}\beta_0 + A_{i,t}X_{i,t}\beta_1 + Z_{i,t}b_{0,i} + A_{i,t}Z_{i,t}b_{1,i} + \epsilon_{i,t+1}.$$

- Want to use historical information ($H_{i,t}$) for the next treatment, not only the current covariate ($X_{i,t}$).

$$H_{i,t} = (X_{i,1}, A_{i,1}, Y_{i,2}, ..., X_{i,t-1}, A_{i,t-1}, Y_{i,t}, X_{i,t}).$$

- $Y_{i,t+1} = f_0(H_{i,t})^\top \beta_0 + A_{i,t} f_1(H_{i,t})^\top \beta_1 + g_0(H_{i,t})^\top b_{0,i} + A_{i,t} g_1(H_{i,t})^\top b_{1,i} + \epsilon_{i,t+1}$.
  $f.$ and $g.$ are some summary functions with a fixed dimension.
- Is there any problem using LMM for this model?
- No, $X_{i,t}$ is often endogenous.

Challenge when using LMM in Precision Medicine, ctd

What is a endogenous covariate?

- Exogenous $= X$ is *not affected* by any variables in the model (e.g. temperature, pre-scheduled / fixed variables, etc)

$$X_{i,t} \perp (H_{i,t-1}, A_{i,t-1}, Y_{i,t})$$

- Endogenous $= X$ is *affected* by some variables in the model (e.g. past outcomes)

$$X_{i,t} = f(H_{i,t-1}, A_{i,t-1}, Y_{i,t})$$

for some function f.

2. QKM's LMM with endogenous covariates

The HeartSteps example

- Want to increase the amount of patients exercise (Y) by push alarm intervention (A)
- We either have push alarm or not every two hours.
- $X_{i,t}$ = number of steps taken during 30 mins before treatment ($A_{i,t}$)
- $Y_{i,t+1}$ = number of steps taken during 30 mins after $A_{i,t}$
- $\underbrace{(X_{i,1}, A_{i,1}, Y_{i,2}), ..., (X_{i,t-1}, A_{i,t-1}, Y_{i,t}), (X_{i,t}}_{=:H_{i,t}}, A_{i,t}, Y_{i,t+1}), ...$

The HeartSteps example

- Want to increase the amount of patients exercise (Y) by push alarm intervention (A)
- We either have push alarm or not every two hours.
- $X_{i,t}$ = number of steps taken during 30 mins before treatment $(A_{i,t})$
- $Y_{i,t+1}$ = number of steps taken during 30 mins after $A_{i,t}$
- $\underbrace{(X_{i,1}, A_{i,1}, Y_{i,2}), ..., (X_{i,t-1}, A_{i,t-1}, Y_{i,t}), (X_{i,t}}_{=:H_{i,t}}, A_{i,t}, Y_{i,t+1}), ...$
- $X_{i,t}$ is obviously endogenous!

QKM's key assumption

The conditional independence assumption

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

QKM's key assumption

The conditional independence assumption

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

$\Rightarrow$ If this assumption (10) holds,
$\hat{\beta}$ estimated by standard LMM packages is a valid MLE.

QKM's LMM - quick proof

$$
\begin{aligned}
&\prod_i p(X_i, A_i, Y_i|\alpha, \beta, \theta, \sigma_\epsilon) \\
&= \prod_i \int p(X_i, A_i, Y_i|b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i) \\
&= \prod_i \{ \int \prod_t \underbrace{p(X_{i,t}|H_{i,t-1}, A_{i,t-1}, Y_{i,t}, b_i)}_{=11A} \underbrace{p(A_{i,t}|H_{i,t}, b_i)}_{=11B} \times \\
&\qquad \underbrace{p(Y_{i,t+1}|H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon)}_{=11C} dF(b_i) \} \\
&= \{ \prod_i \prod_t p(X_{i,t}|H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \, p(A_{i,t}|H_{i,t}) \} \times \\
&\qquad \underbrace{\prod_i \{ \int \prod_t p(Y_{i,t+1}|H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i) \}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon|X,A,Y)}
\end{aligned}
$$

QKM's LMM - quick proof

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) = \underbrace{\{\prod_i \prod_t p(X_{i,t}|H_{i,t-1}, A_{i,t-1}, Y_{i,t})\ p(A_{i,t}|H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times \underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1}|H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon|X,A,Y) = \text{Likelihood of the classic LMM!}}$$

QKM's LMM - quick proof

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) =$$

$$\underbrace{\{\prod_i \prod_t p(X_{i,t} | H_{i,t-1}, A_{i,t-1}, Y_{i,t})\ p(A_{i,t} | H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times$$

$$\underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1} | H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon|X,A,Y) = \text{Likelihood of the classic LMM!}}$$

$$\therefore \arg\max_\xi \mathcal{L}(\underbrace{\alpha, \beta, \theta, \sigma_\epsilon}_{=:\xi} | X, A, Y) = \arg\max_\xi \mathcal{L}_1(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y)$$

QKM's LMM - property

- Once conditional independence assumption holds,
- We can just use standard LMM package.
- Still, β only has conditional-on-random-effects interpretation.

3. Discussion

Discussion

- Is partial likelihood okay to use?
- How to verify the conditional independence?
- Marginal effects estimation
- Nonlinear models - kernel extension

Discussion - 1. partial likelihood

QKM factored out the first two terms assuming that they do not involve $\xi \equiv (\alpha, \beta, \theta, \sigma_\epsilon)$.

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) = \underbrace{\{\prod_i \prod_t p(X_{i,t} | H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \; p(A_{i,t} | H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times$$

$$\underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1} | H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon|X,A,Y) = \text{Likelihood of the classic LMM!}}$$

Discussion - 1. partial likelihood

QKM factored out the first two terms assuming that they do not involve $\xi \equiv (\alpha, \beta, \theta, \sigma_\epsilon)$.

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) = \underbrace{\{\prod_i \prod_t p(X_{i,t} | H_{i,t-1}, A_{i,t-1}, Y_{i,t})\ p(A_{i,t} | H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times \underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1} | H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon | X,A,Y) = \text{Likelihood of the classic LMM!}}$$

However, $X_{i,t}$ might have some information about ξ.

Discussion - 1. partial likelihood

QKM factored out the first two terms assuming that they do not involve $\xi \equiv (\alpha, \beta, \theta, \sigma_\epsilon)$.

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) = \underbrace{\{\prod_i \prod_t p(X_{i,t} | H_{i,t-1}, A_{i,t-1}, Y_{i,t})\ p(A_{i,t} | H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times \underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1} | H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon | X,A,Y) = \text{Likelihood of the classic LMM!}}$$

However, $X_{i,t}$ might have some information about ξ.

$\beta_1 : A_{i,t} \to Y_{i,t+1} \urcorner$ "treatment effect"

$\delta_1 : A_{i,t} \to \quad \to \quad \to X_{i,t+1}$ "delayed treatment effect"

Discussion - 1. partial likelihood

QKM factored out the first two terms assuming that they do not involve $\xi \equiv (\alpha, \beta, \theta, \sigma_\epsilon)$.

$$\mathcal{L}(\alpha, \beta, \theta, \sigma_\epsilon | X, A, Y) = \underbrace{\{\prod_i \prod_t p(X_{i,t} | H_{i,t-1}, A_{i,t-1}, Y_{i,t})\ p(A_{i,t} | H_{i,t})\}}_{\text{Not involving } (\alpha,\beta,\theta,\sigma_\epsilon)} \times \underbrace{\prod_i \{\int \prod_t p(Y_{i,t+1} | H_{i,t}, A_{i,t}, b_i; \alpha, \beta, \theta, \sigma_\epsilon) dF(b_i)\}}_{\mathcal{L}_1(\alpha,\beta,\theta,\sigma_\epsilon | X,A,Y) = \text{Likelihood of the classic LMM!}}$$

However, $X_{i,t}$ might have some information about ξ.

$\beta_1 : A_{i,t} \to Y_{i,t+1}$ ⌝ "treatment effect"

$\delta_1 : A_{i,t} \to \quad \to \quad \to X_{i,t+1}$ "delayed treatment effect"

β_1 and δ_1 may not be orthogonal. So omitting the two terms may cause efficiency loss!

Discussion - 2. verifying the conditional independence

How to verify (10) using data?

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

Discussion - 2. verifying the conditional independence

How to verify (10) using data?

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

- It is not testable without further assumption.

Discussion - 2. verifying the conditional independence

How to verify (10) using data?

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

- It is not testable without further assumption.
- QKM suggests using the domain knowledge to judge independence.

Discussion - 2. verifying the conditional independence

How to verify (10) using data?

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

- It is not testable without further assumption.
- QKM suggests using the domain knowledge to judge independence.
- When $H_{i,t-1}$ contains enough information, additionally having b_i may not help predicting X_{it}. Thus, it is likely conditionally independent.

Discussion - 2. verifying the conditional independence

How to verify (10) using data?

$$X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t} \tag{10}$$

- It is not testable without further assumption.
- QKM suggests using the domain knowledge to judge independence.
- When $H_{i,t-1}$ contains enough information, additionally having b_i may not help predicting X_{it}. Thus, it is likely conditionally independent.
  $X_{i,t+1} = (f_1(\mathbf{H}_{i,t}), A_{i,t}, Y_{i,t+1})^T \gamma_1 + \eta_{i,t+1}$ v.s.
  $X_{i,t+1} = (f_1(\mathbf{H}_{i,t}), A_{i,t}, Y_{i,t+1})^T \gamma_1 + b_i^T \gamma_2 + \eta_{i,t+1}$

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t})$ $\forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \;\; \forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.
- Conditional Distance Independence Test (CDIT) (Wang et al., 2015)

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t})$ $\forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.
- Conditional Distance Independence Test (CDIT) (Wang et al., 2015)

We still have many degrees of freedom.

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \;\; \forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.
- Conditional Distance Independence Test (CDIT) (Wang et al., 2015)

We still have many degrees of freedom.

- Choice of d, the window width:
  Wider d brings curse of dimensionality
  narrower d brings false positives (undue dependence might appear).

Discussion - 2. verifying the conditional independence, ctd

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \ \forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.
- Conditional Distance Independence Test (CDIT) (Wang et al., 2015)

We still have many degrees of freedom.

- Choice of d, the window width:
  Wider d brings curse of dimensionality
  narrower d brings false positives (undue dependence might appear).
- For what set of time points $\mathcal{T}$ to test:
  single test: May not guarantee the results hold for all time points.
  testing on every other r time points.

Develop an ad-hoc test.

- Instead of testing
  $X_{it} \perp (b_{0i}, b_{1i}) | H_{i,t-1}, A_{i,t-1}, Y_{i,t}$ for all of $t = 1, 2, ..., T$,
- we test
  $X_{it} \perp (\hat{b}_{0i}, \hat{b}_{1i}) | s_d(H_{i,t-1}, A_{i,t-1}, Y_{i,t}) \;\; \forall t \in \mathcal{T}$,
  where s_d is a d-dimensional summary function.
- Conditional Distance Independence Test (CDIT) (Wang et al., 2015)

We still have many degrees of freedom.

- Choice of d, the window width:
  Wider d brings curse of dimensionality
  narrower d brings false positives (undue dependence might appear).
- For what set of time points $\mathcal{T}$ to test:
  single test: May not guarantee the results hold for all time points.
  testing on every other r time points.
- How to combine the tests:
  Bonferroni, Benjamini Hochberg, or L_2-norm summarization.

Discussion - 3. marginal effects estimation

wish to estimate

$$E(Y_{it+1}|H_{it}, A_{it} = 1) - E(Y_{it+1}|H_{it}, A_{it} = 0)$$

, or $f_1(h_{it})^T\beta + g_1(h_{it})^T E(b_{1i}|H_{it})$.

Discussion - 3. marginal effects estimation

wish to estimate

$$E(Y_{it+1}|H_{it}, A_{it} = 1) - E(Y_{it+1}|H_{it}, A_{it} = 0)$$

, or $f_1(h_{it})^T \beta + g_1(h_{it})^T E(b_{1i}|H_{it})$.

- One possibility is to posit a linear model:
  $E(\hat{b}_{1ik}|H_{it}) = s_d(H_{it})^T \gamma_k, \qquad k = 1, 2, ..., K, \; t = 1, 2, ..., T.$

Discussion - 3. marginal effects estimation

wish to estimate

$$E(Y_{it+1}|H_{it}, A_{it} = 1) - E(Y_{it+1}|H_{it}, A_{it} = 0)$$

, or $f_1(h_{it})^T \beta + g_1(h_{it})^T E(b_{1i}|H_{it})$.

- One possibility is to posit a linear model: $E(\hat{b}_{1ik}|H_{it}) = s_d(H_{it})^T \gamma_k, \qquad k = 1, 2, ..., K,\ t = 1, 2, ..., T.$
- Then do the OLS.

Discussion - 3. marginal effects estimation

wish to estimate

$$E(Y_{it+1}|H_{it}, A_{it} = 1) - E(Y_{it+1}|H_{it}, A_{it} = 0)$$

, or $f_1(h_{it})^T\beta + g_1(h_{it})^T E(b_{1i}|H_{it})$.

- One possibility is to posit a linear model: $E(\hat{b}_{1ik}|H_{it}) = s_d(H_{it})^T\gamma_k, \qquad k = 1, 2, ..., K,\ t = 1, 2, ..., T.$
- Then do the OLS.
- We show $\hat{\gamma}_k$ is consistent.

Discussion - 4. kernel extension

When the true model is not linear, LMM may not be satisfactory.
Want a more flexible model.

Discussion - 4. kernel extension

When the true model is not linear, LMM may not be satisfactory.
Want a more flexible model.

- QKM's LMM can be naturally extended to non-linear one using kernels.

Discussion - 4. kernel extension

When the true model is not linear, LMM may not be satisfactory.
Want a more flexible model.

- QKM's LMM can be naturally extended to non-linear one using kernels.
- We use the Gaussian kernel width bandwidth γ. By replacing the predictor with the kernel matrix and having an

Discussion - 4. kernel extension

When the true model is not linear, LMM may not be satisfactory.
Want a more flexible model.

- QKM's LMM can be naturally extended to non-linear one using kernels.
- We use the Gaussian kernel width bandwidth γ. By replacing the predictor with the kernel matrix and having an
- L_2-penalty term parametrized by λ, the model becomes a Bayesian LMM. Standard software can be used.

Discussion - 4. kernel extension

When the true model is not linear, LMM may not be satisfactory.
Want a more flexible model.

- QKM's LMM can be naturally extended to non-linear one using kernels.
- We use the Gaussian kernel width bandwidth γ. By replacing the predictor with the kernel matrix and having an
- L_2-penalty term parametrized by λ, the model becomes a Bayesian LMM. Standard software can be used.
- (γ, λ) can be tuned using cross-validation.

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.
- But you have to make sure $X \perp b$ given history.

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.
- But you have to make sure $X \perp b$ given history.
- When the treatment effects on X contains some information on β, the partial likelihood might be inefficient

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.
- But you have to make sure $X \perp b$ given history.
- When the treatment effects on X contains some information on β, the partial likelihood might be inefficient
- Conditional independence test can be used as a diagnostic measures.

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.
- But you have to make sure $X \perp b$ given history.
- When the treatment effects on X contains some information on β, the partial likelihood might be inefficient
- Conditional independence test can be used as a diagnostic measures.
- When interested in marginal effects, it can be estimated by further positing a linear model.

Take home message

- You can use standard software to fit an LMM even with enogenous covariates.
- But you have to make sure $X \perp b$ given history.
- When the treatment effects on X contains some information on β, the partial likelihood might be inefficient
- Conditional independence test can be used as a diagnostic measures.
- When interested in marginal effects, it can be estimated by further positing a linear model.
- LMM can be extended using a kernel extension.

References I