Time-Dependent Covariates

Outline

- Introduction

- Data Example: the MSCM Study

- Stochastic Covariates

- Lagged Covariates

- Time-Dependent Confounders

Introduction

- One advantage of a longitudinal study: to observe the temporal order of exposure and outcome.

- Issues in observational studies.

  - The lag relationship between exposure and outcome.
  - Endogeneity.

Setup

- A common set of discrete observation times $t = 1, \ldots, T$.

- Y_{it} is the response of subject i at time t.

- X_{it} is a time-varying covariate and $\boldsymbol{z}_i$ a vector of baseline or time-invariant variables.

- We assume that X_{it} is measured immediately *after* Y_{it} and therefore cannot be a (direct) *cause* of the response at time t.

Endogenous and exogenous covariate

- In survival analysis: internal and external covariate

- In econometrics literature: factors within/outside the system.

Definition

A covariate process is **exogenous** with respect to the outcome process if the covariate at time t is conditionally independent of all preceding response measurements, that is,

$$f(x_{it} \mid \mathcal{H}_i^Y(t), \mathcal{H}_i^X(t-1), \boldsymbol{z}_i) = f(x_{it} \mid \mathcal{H}_i^X(t-1), \boldsymbol{z}_i)$$

where

$$\mathcal{H}_i^Y(t) = \{Y_{i1}, Y_{i2}, \ldots, Y_{it}\}$$
$$\mathcal{H}_i^X(t-1) = \{X_{i1}, X_{i2}, \ldots, X_{it-1}\}$$

Otherwise the covariate process is said to be **endogenous**.

Examples

- Scheduled observation time, t_{ij} is a time-dependent covariate and is exogenous.

- In crossover trials, the treatment sequence is prescribed and is exogenous.

- After organ transplantation, the dosage of immunosuppressant medication typically depends on the symptoms of graft rejection and is endogenous.

- In Indonesian Children Health Study, xerophthalmia status is time-varying. We can check for endogeneity empirically by regressing X_{it} on both Y_{it}, Y_{it-1}, …, and X_{it-1}, ….

ICHS: Empirically Checking Endogeneity

```
> xerop <- read.table ("../data/xerop.data",
+                       col.names = c("id", "RI", "intercept", "age",
+                       "xero", "cos.time", "sin.time",
+                       "sex", "height.age", "stunted", "time",
+                       "base.age", "season", "time.time"))
> xeropw <- reshape (xerop, direction = "wide",
+                    v.names = c("RI", "xero", "age",
+                    "height.age"),
+                    idvar = c("id", "sex"), timevar = "time",
+                    drop = c("intercept", "cos.time", "sin.time",
+                    "stunted", "base.age", "season", "time.time"))
> summary (glm (xero.2 ~ xero.1 + RI.1, family = "binomial",
+                data = xeropw))
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.3159     0.4208  -7.879 3.29e-15 ***
xero.1        3.4787     0.7798   4.461 8.16e-06 ***
RI.1          0.5683     0.9083   0.626    0.532
> summary (glm (xero.3 ~ xero.2 + RI.2 + RI.1,
+                family = "binomial", data = xeropw))
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.4294     0.5093  -6.733 1.66e-11 ***
xero.2        3.9255     1.0062   3.901 9.57e-05 ***
RI.2        -14.1367  1769.2577  -0.008   0.9936
RI.1          1.6184     0.8729   1.854   0.0637 .
```

Exogeneity

- The assumption of exogeneity allows the factorization

$$
f(\boldsymbol{x}_i, \boldsymbol{y}_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}) =
$$

$$
= \left[\prod_{t=1}^{T} f(y_{it} \mid \mathcal{H}_i^Y(t-1), \mathcal{H}_i^X(t-1), \boldsymbol{z}_i; \boldsymbol{\theta}) \times \prod_{t=1}^{T} f(x_{it} \mid \mathcal{H}_i^X(t-1), \boldsymbol{z}_i; \boldsymbol{\theta}) \right]
$$

$$
= \mathcal{L}_Y(\boldsymbol{\theta}) \times \mathcal{L}_X(\boldsymbol{\theta})
$$

- We further assume that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are variation independent, that is, $(\boldsymbol{\theta}_1 \in \Theta_1) \times (\boldsymbol{\theta}_2 \in \Theta_2)$. Then the process X_{it} is called *strongly exogenous* for the parameter of interest $\boldsymbol{\theta}_1$ if

$$
f(\boldsymbol{x}_i, \boldsymbol{y}_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}) = \mathcal{L}_Y(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \times \mathcal{L}_X(\boldsymbol{\theta}_2).
$$

Inference for $\boldsymbol{\theta}_1$ can be based on the likelihood conditional on $\boldsymbol{x}_i$ without loss of information, and therefore we do not need specify a model for the process X_{it}.

- Exogeneity also implies that

$$\mathrm{E}(Y_{it} \mid x_{i1}, \ldots, x_{iT}, \boldsymbol{z}_i) = \mathrm{E}(Y_{it} \mid x_{i1}, \ldots, x_{it-1}, \boldsymbol{z}_i).$$
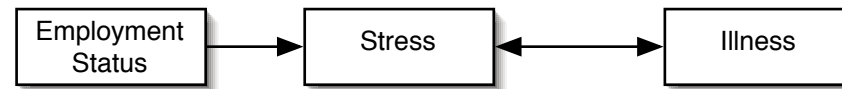
In fact, exogeneity implies a stronger conditional independence condition

$$Y_{it} \perp \{X_{it}, \ldots, X_{iT}\} \mid \mathcal{H}_i^X(t-1).$$

- When a covariate is endogenous, we need be careful in choosing meaningful target of inference and valid methods of estimation.

The MSCM Study

- An observational study of 167 preschool children aging 18 months to 5 years attending an inner-city pediatric clinic.

- Baseline covariates include employment status of the mothers. Four weeks of follow-up with daily measures of maternal stress and child illness.

```
┌────────────┐       ┌────────┐       ┌──────────┐
│ Employment │  ──▶  │ Stress │  ◀──▶ │ Illness  │
│   Status   │       │        │       │          │
└────────────┘       └────────┘       └──────────┘
```

- Questions of interest:

  1. Is there an association between maternal employment and stress?

  2. Is there an association between maternal employment and child illness?

  3. Is there evidence that maternal stress **_causes_** child illness?

- Stress may be in the causal pathway from employment to child illness so we do not adjust for illness in addressing Question 1. No adjustment for stress in Question 2.

Child Illness
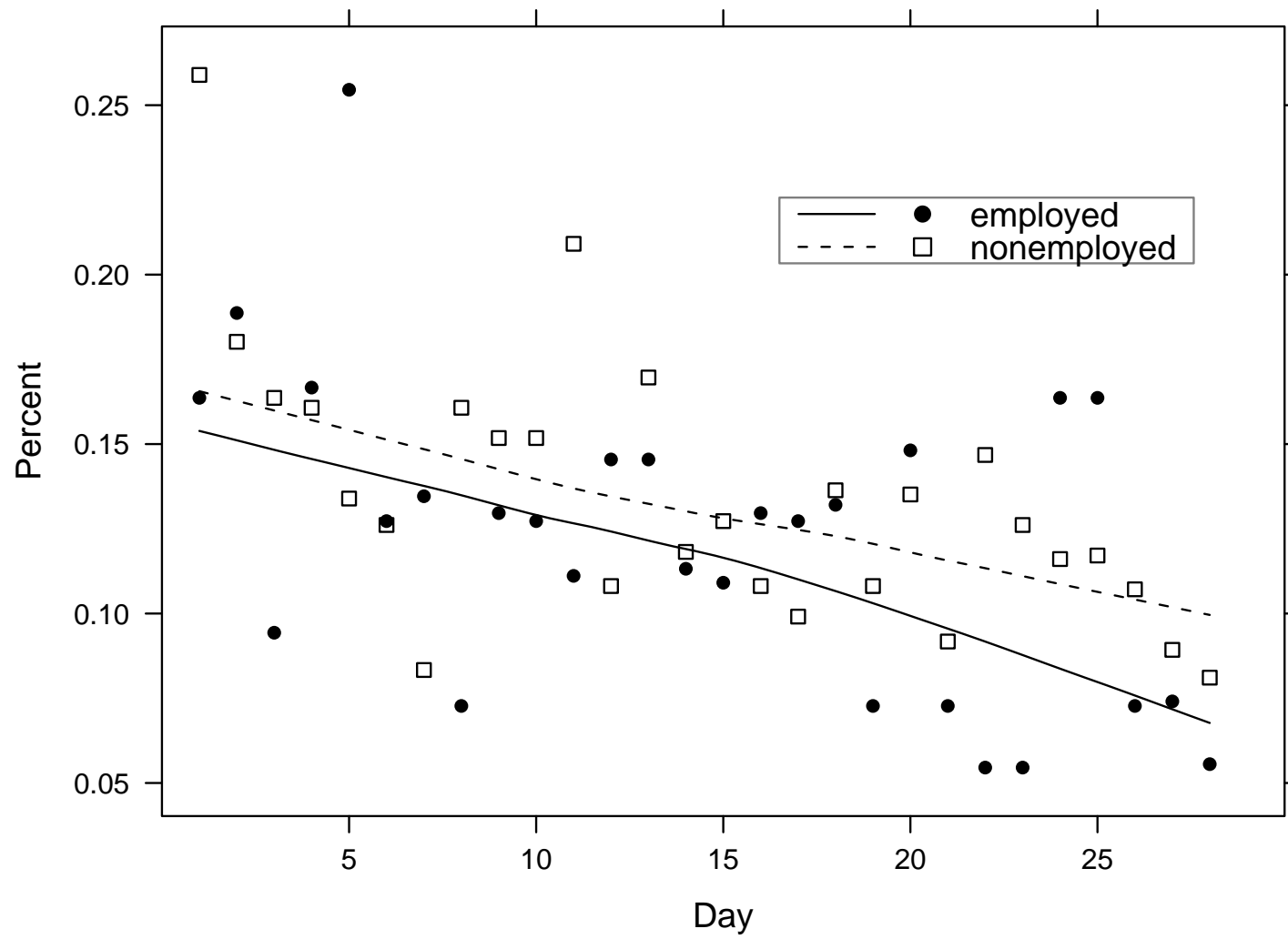
Maternal Stress
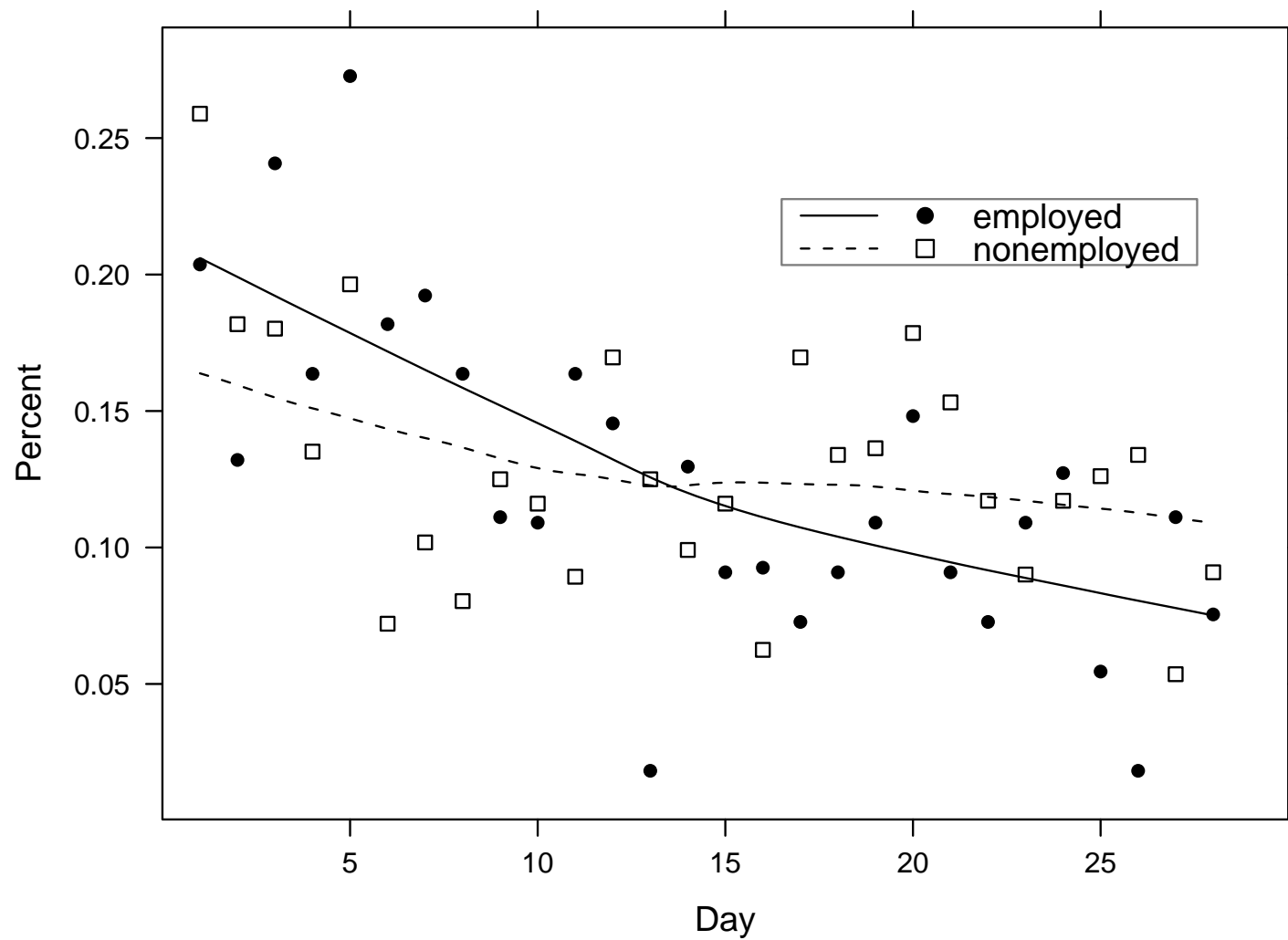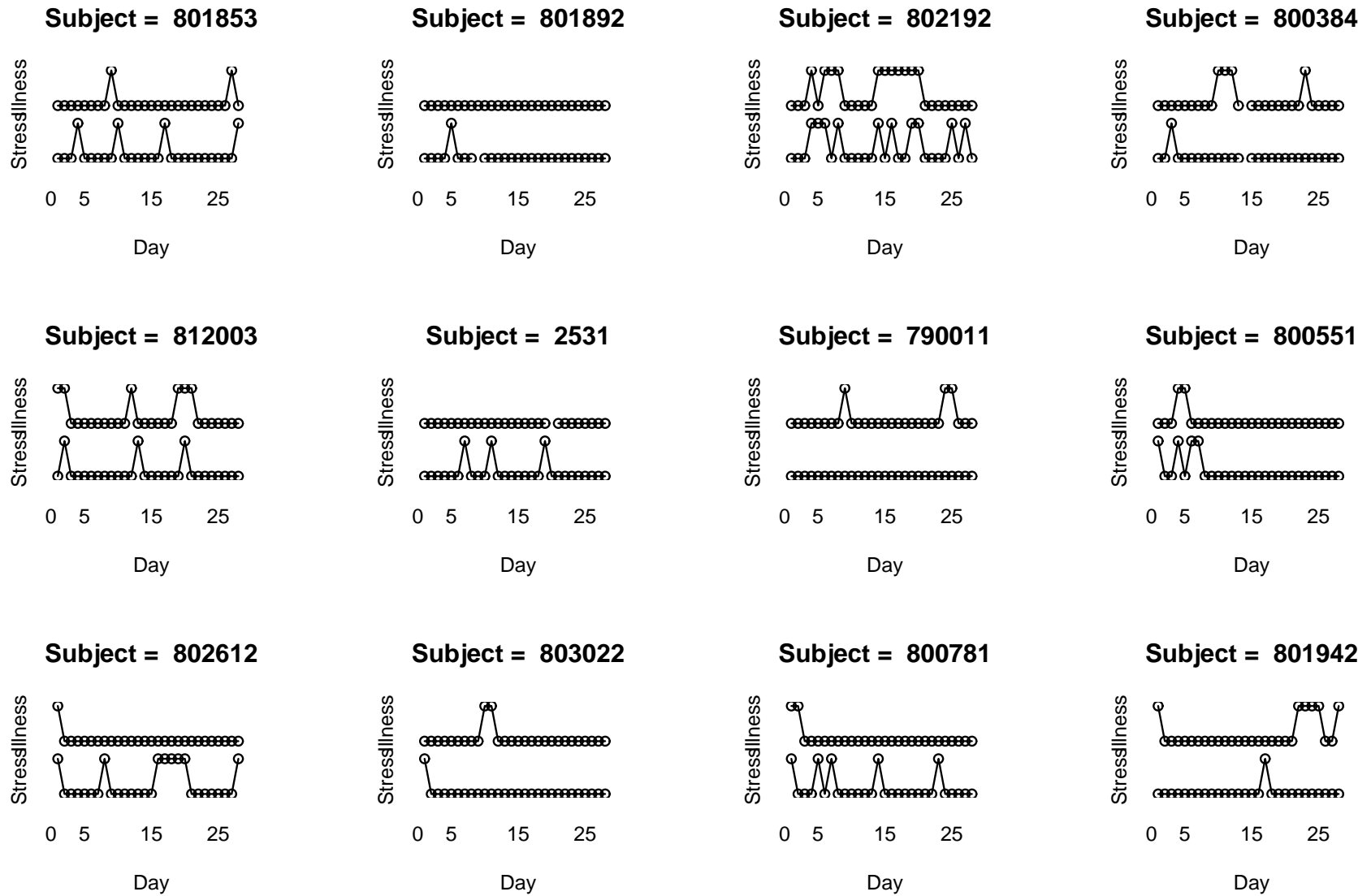
• Several aspects of Question 3, i.e. causal effect of stress on child illness:

1. What is the cross-sectional association between stress on day t and illness on day t? (marginal association)

2. Does illness at day t depend on prior stress measured on day $(t - k)$?

3. What are the factors that influence maternal stress on day t? Does child illness on day $(t - k)$ for $k = 0, 1, 2, \ldots$ predict maternal stress? (endogeneity)

Stochastic Covariates: Full and Partly Conditional Means

- For a stochastic covariate process X_{it} and an outcome process Y_{it}, we may be interested in:

  Concurrent association: $\mathrm{E}(Y_{it} \mid X_{it})$

  Lagged association: $\mathrm{E}(Y_{it} \mid X_{it-k}$ for some $k > 0)$

  Exposure history: $\mathrm{E}(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{it-1})$

  Cumulative exposure: $\mathrm{E}(Y_{it} \mid X_{it}^* = \sum_{s<t} X_{is})$

  Entire process: $\mathrm{E}(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{iT})$

- Define $\mathrm{E}(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{iT})$ as the full covariate conditional mean and

  $\mathrm{E}(Y_{it} \mid \mathrm{subset}\{X_{i1}, X_{i2}, \ldots, X_{iT}\})$ as a partly conditional mean, e.g. the cross-sectional mean $\mathrm{E}(Y_{it} \mid X_{it})$.

- Under the assumption of exogeneity and a finite covariate lag, the partly conditional mean

  $\mathrm{E}(Y_{it} \mid X_{it-1}, X_{it-2}, \ldots, X_{it-k})$ may equal the full conditional mean.

- When the covariate process is *endogenous*,

$$\mathrm{E}(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{iT}) \neq \mathrm{E}(Y_{it} \mid X_{i1}, X_{i2}, \ldots, X_{it-1}),$$

  the full conditional mean is not equal to the scientifically desired partly conditional mean.

- We need first identify which conditional mean is of interest, then identify valid and efficient estimation

  methods.

Marginal Model for Cross-Sectional Association and Estimation

Consider the marginal model for the cross-sectional mean $\mu_{ij} = E[Y_{ij}|X_{ij}]$. For simplicity, we assume that X is a scalar.

- The GEE estimator of the regression parameter $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{S_\beta}(\boldsymbol{\beta}, \boldsymbol{W}) = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}.$$

- The GEE estimator is *consistent* if the estimating function is *unbiased*:

$$\mathrm{E}\left(\boldsymbol{S_\beta}(\boldsymbol{\beta}, \boldsymbol{W})\right) = \boldsymbol{0}.$$

- **When is $\boldsymbol{S}$ unbiased?**

$$S_{\beta_X}(\boldsymbol{\beta}, \boldsymbol{W})$$

$$= \sum_{i=1}^{m} \left[\begin{pmatrix} x_{i1} \\ \vdots \\ x_{in} \end{pmatrix}^T \begin{pmatrix} w_{i11}^* & \cdots & w_{i1n}^* \\ w_{i21}^* & \cdots & w_{i2n}^* \\ \vdots & \ddots & \vdots \\ w_{in1}^* & \cdots & w_{inn}^* \end{pmatrix} \begin{pmatrix} Y_{i1} - \mu_{i1} \\ \vdots \\ Y_{in} - \mu_{in} \end{pmatrix} \right]$$

$$= \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} \sum_{k=1}^{n_i} x_{ij} w_{ijk}^* (Y_{ik} - \mu_{ik}) \right]$$

where

$$\mu_{ik} = \mathrm{E}[Y_{ik}|X_{ik}],$$

$$w_{ijk}^* = \left(\frac{\partial \eta_{ij}}{\partial \mu_{ij}}\right)^{-1} w_{ijk},$$

and w_{ijk} is the (j, k) element of the weight matrix $\boldsymbol{W}_i$.

1. Consider the expectation of

$$\mathrm{E}\left[x_{ij} w_{ijk}^*(Y_{ik} - \mu_{ik})\right] = \mathrm{E}\left\{\mathrm{E}\left[x_{ij} w_{ijk}^*(Y_{ik} - \mu_{ik}) \mid \boldsymbol{X}_i\right]\right\}$$
$$= \mathrm{E}\left\{x_{ij} w_{ijk}^* \left[\mathrm{E}\left(Y_{ik} \mid \boldsymbol{X}_i\right) - \mu_{ik}\right]\right\}$$

The EE is unbiased if the **full covariate conditional mean (FCCM)** assumption is satisfied, that is,

$$\mathrm{E}\left(Y_{ik} \mid \boldsymbol{X}_i\right) = \mu_{ik} \equiv \mathrm{E}(Y_{ik} \mid X_{ik}),$$

where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iT})$
(Pepe and Anderson, 1994)

On the other hand, if FCCM does not hold, the estimating function will likely be biased and result in inconsistent estimates for the cross-sectional mean structure.

2. However, if a diagonal weight matrix is used, then

$$
S_{\beta_X}(\boldsymbol{\beta}, \boldsymbol{W})
$$

$$
= \sum_{i=1}^{m} \left[\begin{pmatrix} x_{i1} \\ \vdots \\ x_{in} \end{pmatrix}^T \begin{pmatrix} w_{i11}^* & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{inn}^* \end{pmatrix} \begin{pmatrix} Y_{i1} - \mu_{i1} \\ \vdots \\ Y_{in} - \mu_{in} \end{pmatrix} \right]
$$

$$
= \sum_{i=1}^{m} \left[\sum_{j=1}^{n} x_{ij} w_{ijj}^* (Y_{ij} - \mu_{ij}) \right]
$$

We have

$$
\mathrm{E}\left[x_{ij} w_{ijj}^* (Y_{ij} - \mu_{ij}) \right] = \mathrm{E}\left\{ \mathrm{E}\left(x_{ij} w_{ijj}^* (Y_{ij} - \mu_{ij}) \mid X_{ij} \right) \right\}
$$
$$
= \mathrm{E}\left\{ x_{ij} w_{ijj}^* \left[\mathrm{E}\left[Y_{ij} \mid X_{ij} \right] - \mu_{ij} \right] \right\}
$$

which is always unbiased (FCCM is not required).

A simulation illustration

Illustrate the FCCM assumption and the failure of methods that use non-diagonal covariance weighting.

- Model assumptions

$$
\begin{aligned}
Y_{it} &= \gamma_0 + \gamma_1 X_{it} + \gamma_2 X_{it-1} + b_i + e_{it}, \\
X_{it} &= \rho X_{it-1} + \epsilon_{it}, \\
b_i, e_{it}, \epsilon_{it} &\sim \quad \text{mutually independent mean zero with}
\end{aligned}
$$

$b_i \sim N(0,1), e_{it} \sim N(0,1), \epsilon_0 \sim N(0,1) \text{ and } \epsilon_{it} \sim N(0, 1 - \rho^2).$

- This model implies

$$
\begin{aligned}
E(Y_{it}|X_{i1}, \ldots, X_{in_i}) &= \gamma_0 + \gamma_1 X_{it} + \gamma_2 X_{it-1}, \\
E(Y_{it}|X_{it}) &= \beta_0 + \beta_1 X_{it},
\end{aligned}
$$

where $\beta_0 = \gamma_0$ and $\beta_1 = \gamma_1 + \gamma_2/\rho.$

Table 1: Average estimates of β_1
in the models with $\gamma_0 = 0, \gamma_1 = 1$, and $\gamma_2 = 1$ for different values of the covariate auto-correlation.

| | $\rho = 0.9$ | $\rho = 0.7$ | $\rho = 0.5$ | $\rho = 0.3$ | $\rho = 0.1$ |
	$\beta_1 = 1.9$	$\beta_1 = 1.7$	$\beta_1 = 1.5$	$\beta = 1.3$	$\beta_1 = 1.1$
Independence	1.90	1.70	1.50	1.30	1.10
Exchangeable	1.73	1.51	1.33	1.19	1.01
AR(1)	1.73	1.36	1.11	0.89	0.74

- Under these assumptions,

$$E[x_{it-1} w^*_{it-1,t}(Y_{it} - \beta_0 - \beta_1 X_{it})] = w^*_{it-1,t} \cdot \gamma_2 \cdot (1 - \rho^2).$$

  – Bias if the covariates are time varying ($\rho \neq 1$), Y_{it} is predicted by $X_{it-1}(\gamma_2 \neq 0)$, and a non-diagonal weight matrix is used ($w^*_{it-1,t} \neq 0$).

- **Conclusion**: the GEE estimator of $\boldsymbol{\beta}$ is consistent if

  1. The full covariate conditional mean (FCCM) assumption holds; or

  2. A working independence weight matrix is used.

- We can always use working independence GEE to model $E(Y_{it} \mid X_{it})$. However we can not interpret the marginal association as causal effects. It is equally plausible that stress causes illness or that illness causes stress. In order to infer cause we need to address the temporal ordering of exposure and outcome.

MSCM study

Analysis of marginal association for MSCM study (DHLZ Table 12.4).

```
> m.ind <- gee (illness ~ stress + week + married + employed + chlth
+               + mhlth + race + education + housize,
+               data = mscm, id = id, family = binomial)


> m.ex <- gee (illness ~ stress + week + married + employed + chlth
+               + mhlth + race + education + housize,
+               data = mscm, id = id, corstr = "exchangeable",
+               family = binomial)


> m.ar1 <- gee (illness ~ stress + week + married + employed + chlth
+               + mhlth + race + education + housize,
+               data = mscm, id = id, corstr = "AR-M", Mv = 1,
+               family = binomial)



> res <- lapply (list (independent = m.ind,
+                      exchangeable = m.ex, ar1 = m.ar1),
+                function (x) (coef (summary (x))[2:3,c(1,4:5)]))
```

```
> options (digits  = 3)
> res
$independent
       Estimate Robust S.E.  Robust z
stress    0.668      0.1394      4.79
week     -0.178      0.0545     -3.26


$exchangeable
       Estimate Robust S.E.  Robust z
stress    0.532      0.1363      3.91
week     -0.182      0.0545     -3.34


$ar1
       Estimate Robust S.E.  Robust z
stress    0.379      0.1220      3.10
week     -0.206      0.0539     -3.82


> m.ex$working.correlation[1,2]
[1] 0.0615


> m.ar1$working.correlation[1,2]
[1] 0.39
```

Lagged Covariates

Scenarios of using lagged covariates:

- A single lagged covariate, X_{it-k}.

- In smoking studies, use cumulative exposure, $X_{it}^* = \sum_{s<t} X_{is}$, pack-years, as the summary of exposure history.

- Use the entire covariate history $X_{i1}, \cdots, X_{it}$ as potential predictors for Y_{it}.

- Use only a subset of recent covariates, $X_{it-1}, \cdots, X_{it-k}$ when an acute effect is assumed.

In the latter two cases, the lagged covariates can be highly correlated.

A Single Lagged Covariate

In a *partly conditional* GLM

$$g(\mathrm{E}(Y_{it} \mid X_{is}, \boldsymbol{Z}_i)) = \beta_0(t, s) + X_{is}\beta_1(t, s) + \boldsymbol{Z}_i^T\boldsymbol{\beta}_2(t, s).$$

- It is called "partly" because only one covariate time is included as predictor.

- β_1 may depend on both the response time t and the covariate time s.

- Often β_1 is assumed to be a function of only lag, $(t - s)$. Given a particular functional form for $\beta_1(t, s)$, we can use GEE with working independence to estimate the parameters.

Example: MSCM data

$$\mathrm{logit}(\mathrm{E}(Y_{it} \mid X_{it-k}, \boldsymbol{Z}_i)) = \beta_0(k) + \beta_1(k)X_{it-k} + \boldsymbol{Z}_i^T\beta_2(k)$$

- Fitting separate models for $k = 1, 2, \ldots, 7$.

- Assume β_2 constant for k, and some parametric function (of $k = 1, \ldots, 29$), for $\beta_0(k)$ and $\beta_1(k)$.

```
#DHLZ Figures 12.3
> library(magic)
> library(abind)
> mscm$s7 <- shift (mscm$s6, id = mscm$id)
> il.lags <- list ()
> for (ii in paste ("s", 1:7, sep = "")) {
+     il.lags[[ii]] <- gee (illness ~ mscm[,ii] + week + married +
+                           employed + chlth + mhlth + race +
+                           education + housize, data = mscm,
+                           id = id, family = binomial)
+ }
> il.lags.coef <- do.call ("rbind",
+                          lapply (il.lags, function (x) {
+                              coef (summary (x))[2,c(1,4)]
+                          }))
> il.lags.coef <- data.frame (il.lags.coef)
> names (il.lags.coef) <- c("beta", "se")
> with (il.lags.coef, {
+     plot (beta, xlim = c(0, 30), ylim = c(-0.6, 0.8),
+           xlab = "Time lag (days)",
+           ylab = "Coefficient (log odds ratio)",
+           main = "MSCM Study: Lag coefficient function")
+     abline (h = 0, lty = 2)
+     errbar (1:7, beta, beta - se * 1.96, beta + se * 1.96,
+             add = TRUE)
+ })
```

MSCM Study: Lag coefficient function
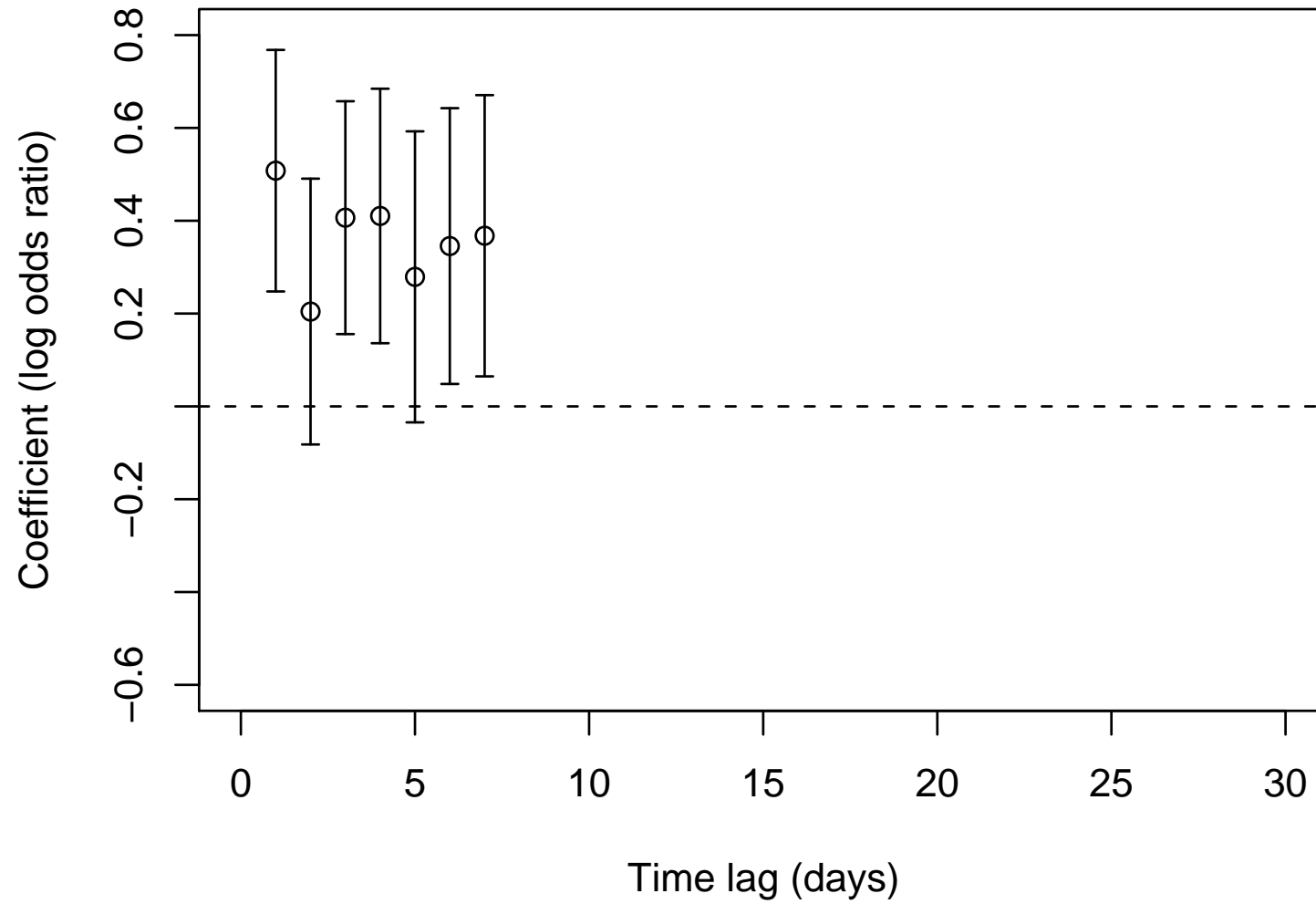
Multiple Lagged Covariates

Sometimes we may need models with multiple lagged variables:

$$g(\mathrm{E}(Y_{it} \mid \mathcal{H}_{it-1}^{X}, \boldsymbol{Z}_i)) = \beta_0 + \beta_1 X_{it-1} + \cdots + \beta_L X_{it-L} + \boldsymbol{Z}_i^T \boldsymbol{\alpha},$$

for a finite lag L. The above model may be over-specified and lead to estimates of lagged covariate effects that are difficult to interpret.

We can further assume that the lagged coefficents $(\beta_1, \beta_2, \ldots, \beta_L)$ follow a lower order smooth parametric function (**distributed lag models**):

$$\beta_j = \gamma_0 + \gamma_1 B_1(j) + \cdots + \gamma_p B_p(j)$$

where $\boldsymbol{B}(j)$ is an appropriate basis vector. For polynomial model $B_l(j) = j^l, l = 0, \cdots, p$.

- We have

$$\boldsymbol{\beta}_{(L+1)\times 1} = B_{(p+1)\times(L+1)}^{T} \boldsymbol{\gamma}_{(p+1)\times 1}.$$

and the regression can be done using transformed variables $X_i B^T$

$$g(\boldsymbol{\mu}_i) = X_i \boldsymbol{\beta} + Z_i \boldsymbol{\alpha} = (X_i B^T) \boldsymbol{\gamma} + Z_i \boldsymbol{\alpha}.$$

- The coefficient model allows parsimonious modeling.

- We need choose the appropriate L (number of lagged covariates) and p (order of the coefficient model).

MSCM data (DHLZ Figures 12.4 and Table 12.5)

Saturated model

```
> iT5.sat <- gee (illness ~ s1 + s2 + s3 + s4 + s5 + s6 + s7 +
+               week + married + employed +
+               chlth + mhlth + race + education + housize,
+               data = mscm, id = id,
+               family = binomial)
> iT5sat.e <- coef (summary (iT5.sat))[2:8, c(1,4)]
> colnames (iT5sat.e) <- c("Beta", "SE")
> rownames (iT5sat.e) <- c("s1","s2","s3","s4","s5","s6","s7")
> iT5sat.e
          Beta         SE
s1  0.34329239 0.1561614
s2 -0.05151545 0.1509723
s3  0.17934857 0.1363666
s4  0.26664393 0.1373006
s5  0.23557409 0.1463803
s6  0.20839598 0.1421672
s7  0.26424049 0.1367730
```

Natural spline with 4 degrees of freedom

```
B <- ns (1:7, knots = c(3, 5), intercept = TRUE)
Xstar <- as.matrix (mscm[,paste("s",1:7,sep="")]) %*% B
iT5.ns <- gee (illness ~ Xstar +
                week + married + employed +
                chlth + mhlth + race + education + housize,
                data = mscm, id = id,
                family = binomial)
tmp.c <- iT5.ns$coef[2:5]
tmp.v <- iT5.ns$robust[2:5, 2:5]
iT5ns.e <- cbind (B %*% tmp.c,
                sqrt (diag (B %*% tmp.v %*% t(B))))
> colnames (iT5ns.e) <- c("Beta", "SE")
> rownames (iT5ns.e) <- c("s1","s2","s3","s4","s5","s6","s7")
> iT5ns.e
        Beta         SE
s1 0.2517556 0.15824540
s2 0.1450647 0.11694727
s3 0.1088346 0.12277381
s4 0.1808002 0.09622847
s5 0.2677936 0.11855602
s6 0.2741420 0.11505092
s7 0.2250573 0.13545063
```

Monotone model, $\beta_j = \gamma_0 + \gamma_1 \cdot (1/j)$

```
> M <- matrix(c(rep(1,7),1,1/2,1/3,1/4,1/5,1/6,1/7),ncol=2)
> t(M)
     [,1] [,2]      [,3] [,4] [,5]      [,6]      [,7]
[1,]    1  1.0 1.0000000 1.00  1.0 1.0000000 1.0000000
[2,]    1  0.5 0.3333333 0.25  0.2 0.1666667 0.1428571


> Xstar.mono <- as.matrix (mscm[,paste("s",1:7,sep="")]) %*% M
> iT5.mono <- gee (illness ~ Xstar.mono +week + married + employed +
+                  chlth + mhlth + race + education + housize,
+                  data = mscm, id = id,family = binomial)
> tmp.c <- iT5.mono$coef[2:3]
> tmp.v <- iT5.mono$robust[2:3, 2:3]
> iT5mono.e <- cbind (M %*% tmp.c,
+                  sqrt (diag (M %*% tmp.v %*% t(M))))
> colnames (iT5mono.e) <- c("Beta", "SE")
> rownames (iT5mono.e) <- c("s1","s2","s3","s4","s5","s6","s7")
> iT5mono.e
        Beta         SE
s1 0.2224116 0.17224088
s2 0.2110720 0.07948940
s3 0.2072921 0.07137674
s4 0.2054021 0.07558084
s5 0.2042682 0.08047788
s6 0.2035122 0.08455853
s7 0.2029722 0.08781513
```

```
#DHLZ Figures 12.4
iT5sat.e <- data.frame (iT5sat.e)
with (iT5sat.e, {
    plot (Beta, xlim = c(0.5, 7.5), ylim = c(-0.5, 0.8), xaxt="n", pch=1,
          xlab = "Time lag (days)", ylab = "Coefficient (log odds ratio)",
          main = "MSCM Study: Lag coefficient function")
    axis(side=1,at=1:7,lab=c("1","2","3","4","5","6","7"))
    abline (h = 0, lty = 2)
    errbar (1:7, Beta, Beta - SE * 1.96, Beta + SE * 1.96,
            add = TRUE)})
iT5ns.e <- data.frame (iT5ns.e)
with (iT5ns.e, {
    points (1:7+0.1,Beta, pch=12)
    lines (1:7+0.1,Beta, lty = 3)
    abline (h = 0, lty = 2)
    errbar (1:7+0.1, Beta, Beta - SE * 1.96, Beta + SE * 1.96,
            add = TRUE)})
iT5mono.e <- data.frame (iT5mono.e)
with (iT5mono.e, {
    points (1:7+0.2,Beta, pch=6)
    lines (1:7+0.2,Beta, lty = 4)
    abline (h = 0, lty = 2)
    errbar (1:7+0.2, Beta, Beta - SE * 1.96, Beta + SE * 1.96,
            add = TRUE)})
legend(5,-0.2,c("Saturated","Spline","Monotone"),pch=c(1,12,6),cex=1)
```

MSCM Study: Lag coefficient function
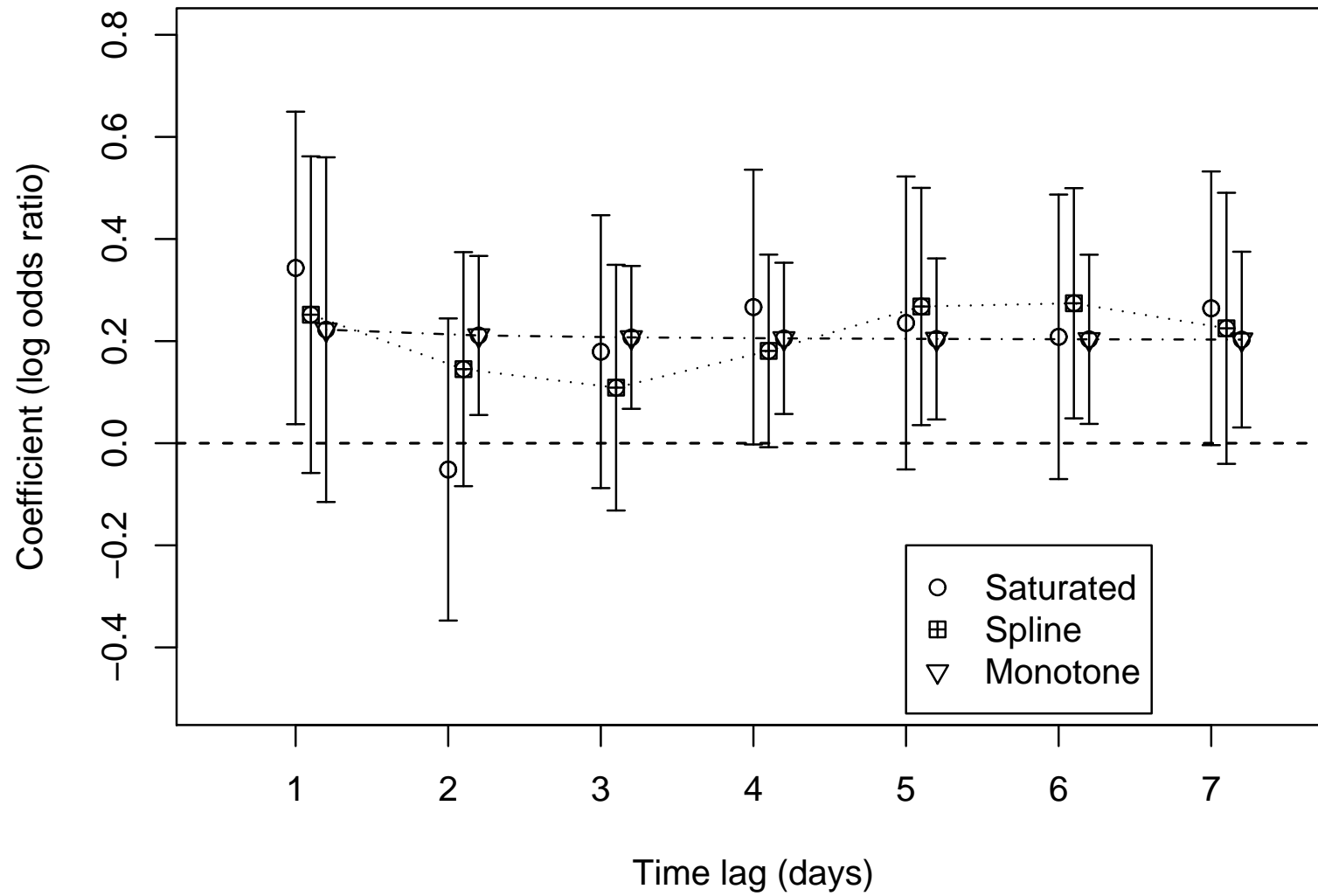
- Each of these models shows an association between maternal stress in the previous 7 days and current child illness.

- We cannot use AIC or BIC to compare the adequacy of different distributed lag models.

- However, we can evaluate the **predictive accuracy** of each model (Harrell et al., 1984).

  1. Delete individual subjects

  2. Refit the model and compare observed and fitted outcome vectors

  3. Use c-index (area under the ROC curve) as a global summary of model accuracy.

- The c-index for the saturated model: 64.1%, the spline model: 63.8%, and the monotone model: 64.2%. They provide almost identical predictive accuracy.

Time-Dependent Confounders

- When a covariate is associated with both the exposure of interest and the outcome of interest:

  – a *confounder* if ignoring it leads to biased exposure effect estimates.

  – an *intermediate variable* if it is in the causal pathway from exposure to outcome.

- We should adjust for the confounders but not the intermediate variables.

- In longitudinal studies, the time-dependent variable can be both a confounder and an intermediate variable.

- MSMC data: endogeneity of stress can be tested by regressing stress at time t on stress at time $t-1$, $t-2$, ..., and on child illness at time t and $t-1$ (DHLZ Table 12.7).
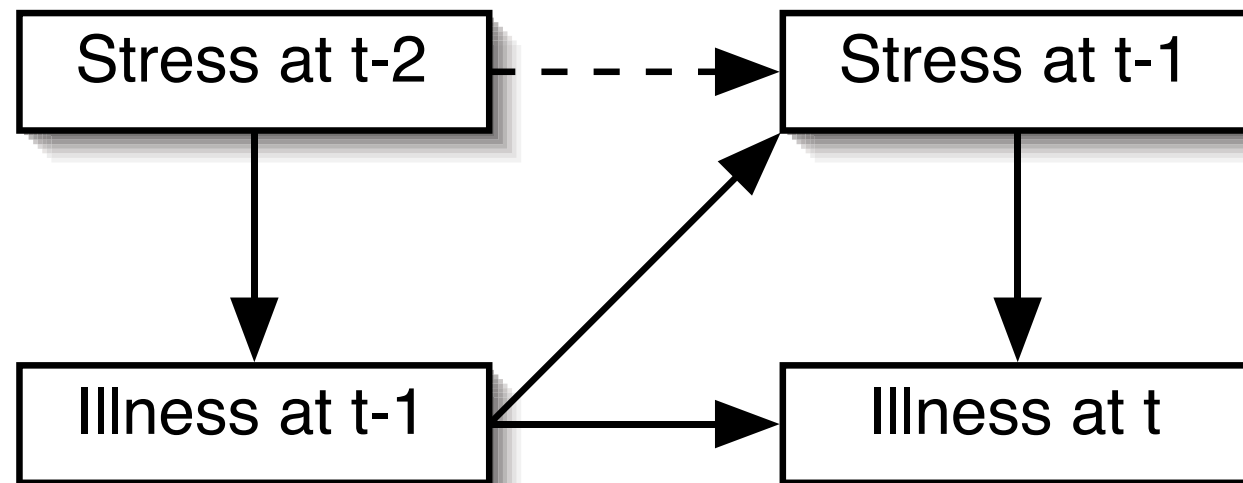
Causal Inference

- **Counterfactual/potential outcome**

  – Define $Y_{it}^{(\boldsymbol{x}_t)}$ as the outcome for subject i at time t that would be observed if a given covariate sequence $\boldsymbol{x}_t = (x_1, x_2, \ldots, x_{t-1})$ was in force.

  – For a given subject we can only observe a single outcome at time t for a specific covariate sequence $\boldsymbol{x}_t^*$. All other possible outcomes $Y_{it}^{(\boldsymbol{x}_t)}$ for $\boldsymbol{x}_t \neq \boldsymbol{x}_t^*$ are not observed.

- The **causal effect** of $\boldsymbol{x}_t = (1, 1, \ldots, 1)$ versus $\boldsymbol{x}_t = (0, 0, \ldots, 0)$ at time t is defined as

$$\delta_t = \mathrm{E}\left[Y_{it}^{(\mathbf{1})} - Y_{it}^{(\mathbf{0})}\right] = \mathrm{E}\left[Y_{it}^{(\mathbf{1})}\right] - \mathrm{E}\left[Y_{it}^{(\mathbf{0})}\right].$$

  – For each i, at most only one of $Y_{it}^{(\mathbf{0})}$ or $Y_{it}^{(\mathbf{1})}$ is observed, or neither is observed.

  – The causal effect refers to the effect of interventions in the entire population rather than possibly selected, observed subjects.

– For a randomized trial with full compliance, $\mathrm{E}[Y_{it} \mid \boldsymbol{x}_t = \boldsymbol{1}]$ is an unbiased estimate of $\mathrm{E}\left[Y_{it}^{(\boldsymbol{1})}\right]$.

However, in other cases, $\mathrm{E}[Y_{it} \mid \boldsymbol{x}_t = \boldsymbol{1}]$ is not necessarily an unbiased estimate of $\mathrm{E}\left[Y_{it}^{(\boldsymbol{1})}\right]$. Note

that $\mathrm{E}[Y_{it} \mid \boldsymbol{x}_t]$ is the conditional expectation given the observed treatment, while $\mathrm{E}\left[Y_{it}^{(\boldsymbol{1})}\right]$ is the

population average if the treatment is enforced to the whole population.

– With additional baseline covariates $\boldsymbol{Z}$, we define

$$\mu_t^{(\boldsymbol{x}_t)}(\boldsymbol{z}) = \mathrm{E}\left(Y_{it}^{(\boldsymbol{x}_t)} \mid \boldsymbol{Z} = \boldsymbol{z}\right)$$

as the average outcome that would be observed at time t within the sub-population defined by

$\boldsymbol{Z} = \boldsymbol{z}$ if all subjects following the treatment/exposure path $\boldsymbol{x}_t$.

– The causal effect of continuous exposure is defined as

$$\delta_T(\boldsymbol{z}) = \mu_T^{(\boldsymbol{1})}(\boldsymbol{z}) - \mu_T^{(\boldsymbol{0})}(\boldsymbol{z})$$

where T represents the end of the study.

- Robins et al (1999) formalized the definition of **no unmeasured confounders**.

    – In the absence of unmeasured confounders, association implies (defines) causation.

    – Formally, the assumption of "no unmeasured confounders" or **sequential randomization assumption** for longitudinal data is

    $$\left\{ \left(Y_s^{(1)}, Y_s^{(0)} \right); s = t+1, t+2, \dots T \right\} \perp X_t \,|\, \mathcal{H}^X(t-1), \mathcal{H}^Y(t)$$

    – That is, the exposure at time t is independent of any future potential outcomes at time $s > t$, given information through time t.

    – One example where this assumption would be violated: a physician prescribes treatment to the sickest patients.

    – This assumption (and assumptions needed in causal inference in general) are not empirically verifiable.

- Pearl (2000) argued that causal statements require new notation

$$E[Y_2|\text{do}(X_0 = 1, X_1 = 1)]$$

to denote the outcome for the entire population (which is the same as $\mu_2^{(1)}$).

  - His notation emphasizes the fact that we are interested in an average outcome after assignment of

    treatment rather than the average of outcomes in subgroups after simply observing the treatment

    status.

- Estimation approaches for causal inference (mostly due to Jim Robins):

  - g-computation algorithm.

  - Semiparametric g-estimation of structured nested mean models (SNMM).

  - Marginal structural models (MSM) using inverse probability of treatment weights (IPTW).
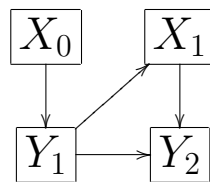
Artificial Example

DHLZ considered an artificial data with two time points and binary covariate X and outcome Y. Data are generated from the following sequential conditional models:

$$\text{logit E}(Y_1 \mid X_0 = x_0) \;=\; -0.5 - 0.5x_0, \tag{1}$$

$$\text{logit E}(X_1 \mid Y_1 = y_1, X_0 = x_0) \;=\; -0.5 + 1.0y_1 \tag{2}$$

$$\text{logit E}(Y_2 \mid X_1 = x_1, X_0 = x_0, Y_1 = y_1) \;=\; -1.0 + 1.5y_1 - 0.5x_1 \tag{3}$$

- Marginal expectations

$$\Pr(Y_1 = 1 \mid x_0 = 0) = \frac{e^{-0.5}}{1 + e^{-0.5}} = 0.38$$

$$\Pr(Y_1 = 1 \mid x_0 = 1) = \frac{e^{-1}}{1 + e^{-1}} = 0.27$$

$$\Pr(X_1 = 1 \mid Y_1 = 0, x_0) = \frac{e^{-0.5}}{1 + e^{-0.5}} = 0.38$$

$$\Pr(X_1 = 1 \mid Y_1 = 1, x_0) = \frac{e^{0.5}}{1 + e^{0.5}} = 0.62$$

$$\Pr(Y_2 = 1 \mid X_0 = 1, X_1 = 1) = 0.30$$

$$\Pr(Y_2 = 1 \mid X_0 = 0, X_1 = 0) = 0.36.$$

The observed odds ratio is

$$\frac{0.30/(1 - 0.30)}{0.36/(1 - 0.36)} = 0.76.$$

- The marginal distribution of Y_2 conditional on (X_0, X_1) is

$$\text{logit E}(Y_2 \mid x_0, x_1) = -0.56 - 0.13x_0 - 0.10x_1 - 0.04x_0 x_1. \tag{4}$$

 – The marginal model above reflects an effect of X_0 in addition to that of X_1 on Y_2.

 – Since Y_1 is correlated with both X_1 and Y_2, the observed marginal effects do not account for the confounder Y_1, and hence do not reflect the causal effect of treatment X_1.

 – At the same time, Y_1 is also an intermediate variable. If we adjust for Y_1 (i.e. model (3)) the effect of X_0 will be masked (no effect).

 – In summary, no standard regression methods can be used to obtain causal statements.

- In order to calculate the causal effects,

  – we can substitute the assignment of treatment X_1, $[X_1|Y_1, X_0]$, with a deterministic assignment mechanism

  – but not alter the response models $\Pr(Y_1|X_0)$ or $\Pr(Y_2|Y_1, X_0, X_1)$.

  – The potential outcomes are (DHLZ Table 12.8)

$$\mu_2^{(1)} = E[Y_2|\text{do}(X_0 = 1, X_1 = 1)] = 0.267$$
$$\mu_2^{(0)} = E[Y_2|\text{do}(X_0 = 0, X_1 = 0)] = 0.402$$

  – The causal odds ratio is 0.54 (a 46% decrease in odds) versus the observed odds ratio of 0.76 (a smaller, 24% decrease in odds).

Table 2: Expected outcomes when treatment is controlled and the causal path leading from Y_1 to X_1 is blocked

---

All subjects $X_0 = X_1 = 1$

X_0						0							1			
n						0							1000			
Y_1			0					1			0				1		
n			0					0			731				269		
X_1		0		1		0		1	0		1		0			1	
n		0		0		0		0	0		731		0			269	
Y_2	0		1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
n	0		0	0	0	0	0	0	0	0	0	598	133	0	0	134	134

$$\mu^{(1)} = (133 + 134)/1000 = 0.267$$

All subjects $X_0 = X_1 = 0$

X_0						0							1			
n						1000							0			
Y_1			0					1			0				1		
n			622					378			0				0		
X_1		0		1		0		1	0		1		0			1	
n		622		0		378		0	0		0		0			0	
Y_2	0		1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
n	455		167	0	0	143	235	0	0	0	0	0	0	0	0	0	0

$$\mu^{(0)} = (167 + 235)/1000 = 0.402$$

---

g-**Computation**

- In this approach, causal effect estimates are obtained from estimates of the observed response transition model under the assumption of no unmeasured confounders.

- The likelihood for a binary response $Y_{i1}, \ldots, Y_{iT}$ and a binary time-dependent covariate $X_{i0}, \ldots, X_{iT-1}$ can be decomposed into

$$
\begin{aligned}
\mathcal{L} &= \prod_{t=1}^{T} \Pr(Y_{it} \,|\, \mathcal{H}_i^Y(t-1), \mathcal{H}_i^X(t-1), \boldsymbol{Z}_i) \times \prod_{t=1}^{T} \Pr(X_{it-1} \,|\, \mathcal{H}_i^Y(t-1), \mathcal{H}_i^X(t-2), \boldsymbol{Z}_i) \\
&= \mathcal{L}_Y \times \mathcal{L}_X
\end{aligned}
$$

- Unknown parameters in the response transition model $\mathcal{L}_Y$ and covariate transition model $\mathcal{L}_X$ can be estimated using maximum likelihood.

- Under the assumption of no unmeasured confounders

$$
\Pr\left(Y_{it}^{(\boldsymbol{x}_t)} = 1 \,|\, \boldsymbol{Y}_{it-1}^{(\boldsymbol{x}_t)}, \boldsymbol{Z}_i\right) = \Pr\left(Y_{it} = 1 \,|\, \mathcal{H}_i^Y(t-1), \mathcal{H}_i^X(t-1) = \boldsymbol{x}_t, \boldsymbol{Z}_i\right).
$$

(*Proposition 2* in Robins et al, 1999)
Hence, we an use the distribution of the observed response at time t among subjects with an observed treatment path and observed response history to estimate the conditional distribution of the counterfactual outcome $Y_{it}^{(\boldsymbol{x}_t)}$.

- We obtain the marginal distribution of the outcome at the final time t, $\Pr\left(Y_{it}^{(\boldsymbol{x}_t)} \mid \boldsymbol{Z}_i\right)$, by using the conditional probabilities to obtain the joint probability for $Y_{i1}^{(\boldsymbol{x}_t)}, \ldots, Y_{it}^{(\boldsymbol{x}_t)}$ and the summing over all possible intermediate paths for the first $t-1$ outcomes.

$$\mu^{(\boldsymbol{x}_t)}(\boldsymbol{z}) = \Pr[Y_{it}^{(\boldsymbol{x}_t)} = 1 | \boldsymbol{Z}_i = \boldsymbol{z}],$$

$$\Pr[Y_{it}^{(\boldsymbol{x}_t)} | \boldsymbol{Z}_i = \boldsymbol{z}]$$

$$= \sum_{\boldsymbol{y}_{t-1}} \Pr[Y_{it}^{(\boldsymbol{x}_t)}, Y_{it-1}^{(\boldsymbol{x}_t)}, \cdots, Y_{i1}^{(\boldsymbol{x}_t)} | \boldsymbol{Z}_i = \boldsymbol{z}]$$

$$= \sum_{\boldsymbol{y}_{t-1}} \prod_{s=1}^{t} \Pr[Y_{is}^{(\boldsymbol{x}_t)} | Y_{is-1}^{(\boldsymbol{x}_t)}, \cdots, Y_{i1}^{(\boldsymbol{x}_t)}; \boldsymbol{Z}_i = \boldsymbol{z}]$$

$$= \sum_{\boldsymbol{y}_{t-1}} \prod_{s=1}^{t} \Pr[Y_{is} | \mathcal{H}^Y(s-1) = \boldsymbol{y}_{s-1}, \mathcal{H}^X(s-1) = \boldsymbol{x}_{s-1}; \boldsymbol{Z}_i = \boldsymbol{z}]$$

- In the artificial data example,

$$\mu_2^{(1)} = \sum_{y_1=0}^{1} \Pr(Y_2 = 1 | X_1 = 1, Y_1 = y_1, X_0 = 1) \cdot \Pr(Y_1 = y_1 | X_0 = 1),$$

$$\mu_2^{(0)} =$$

- In estimating the causal effect, we are controlling the sequence of exposures, but are allowing the intermediate outcome $Y_s, s < t$, to unfold and therefore we capture any indirect effects that are mediated by these earlier outcomes.

- This is a special case of the *g-computation algorithm*. In general, this g-computation formula can be evaluated using Monte Carlo.

- In summary, if we collect adequate covariates Z such that we are willing to assume the no unmeasured confounders condition, then we can use observed data to model the sequence of conditional distribution $Pr[Y_{is} | \mathcal{H}^Y(s-1), \mathcal{H}^X(s-1); Z_i]$ and then use these to evaluate $\mu^{(\boldsymbol{x}_t)}(\boldsymbol{z})$ and hence $\delta_t(\boldsymbol{z})$.

- One advantage of g-computation is that a model for the treatment is not required.

- One limitation is that no direct regression model parameter represents the null hypothesis of no causal effect, thus does not facilitate formal testing.

Inverse Probability of Treatment Weights (IPTW)

- The **marginal structured models** (MSMs) (Robins, 1998; Hernan et al., 2001) allows direct regression modeling of causal effects but requires a model for the covariate process.

- The MSM models the expectation of the counterfactual outcome directly.

$$\mu_t^{(\boldsymbol{x}_t)}(\boldsymbol{z}) = \mathrm{E}\left(Y_{it}^{(\boldsymbol{x}_t)} \mid \boldsymbol{Z}_i = \boldsymbol{z}\right)$$
$$g(\mu_t^{(\boldsymbol{x}_t)}) = \beta_0 + \boldsymbol{\beta}_1^T \boldsymbol{X}_{it}^* + \boldsymbol{\beta}_2^T \boldsymbol{Z}_i$$

where $\boldsymbol{X}_{it}^*$ is a function of the covariate history.

- The parameter $\boldsymbol{\beta}_1$ can be used to quantify and test the causal effect of exposure.

- We cannot use GEE directly because the prior responses are confounder. We also do not want to control for prior responses because they are also intermediate variables.

- Estimation is done using IPTW where weights are used to construct a **pseudo-population** for which the previous outcomes are no longer confounders.

- The **stabilizing weights** are defined by

$$\text{SW}_i(t) = \prod_{s<t} \frac{\Pr\left(X_{is} = x_{is} \mid \mathcal{H}_i^X(s-1), \boldsymbol{Z}_i\right)}{\Pr\left(X_{is} = x_{is} \mid \mathcal{H}_i^Y(s-1), \mathcal{H}_i^X(s-1), \boldsymbol{Z}_i\right)}$$

- In our hypothetical example,

$$\text{SW}(2) = \frac{\Pr(X_1 \mid X_0)}{\Pr(X_1 \mid X_0, Y_1)}.$$

Table 12.10 in DHLZ.

X_0	Y_1	X_1	Y_2	Expected count	Weight	Re-weighted count
1	1	1	1	41.9	0.712	29.8
1	1	1	0	41.9	0.712	29.8
1	1	0	1	31.6	1.474	46.6
1	1	0	0	19.2	1.474	28.3
1	0	1	1	25.2	1.174	29.6
1	0	1	0	112.8	1.174	132.5
1	0	0	1	61.2	0.894	54.7
1	0	0	0	166.3	0.894	148.7
...						

- The weights compare the probability of the treatment received through time $t-1$ conditional only on the treatment histories to the probability of treatment received conditional on both treatment and outcome histories.

- The weights would be 1 if exogeneity is satisfied and also serves as a measure of endogeneity.

- Weights need to be estimated by choosing and fitting models for both the numerator and denominator. Correct modeling for the denominator is needed for consistent parameter estimation, while the model choice for the numerator only impacts efficiency not validity.

- GEE with working independence and the estimated weights $\widehat{\mathrm{SW}}_i(t)$ can be used to estimate $\boldsymbol{\beta}_1$ with empirical sandwich variance estimates.

MSCM Data

- Marginal association model (i.e. the saturated model with multiple lag covariates, DHLZ Table 12.5)

```
> iT5.sat <- gee (illness ~ s1 + s2 + s3 + s4 + s5 + s6 + s7 +
+              week + married + employed +
+              chlth + mhlth + race + education + housize,
+              data = mscm, id = id,
+              family = binomial)
> iT5sat.e <- coef (summary (iT5.sat))[2:8, c(1,4)]
> colnames (iT5sat.e) <- c("Beta", "SE")
> rownames (iT5sat.e) <- c("s1","s2","s3","s4","s5","s6","s7")
> iT5sat.e
         Beta        SE
s1   0.34329239 0.1561614
s2  -0.05151545 0.1509723
s3   0.17934857 0.1363666
s4   0.26664393 0.1373006
s5   0.23557409 0.1463803
s6   0.20839598 0.1421672
s7   0.26424049 0.1367730
```

The marginal effect is

$$\beta_{X_{t-1}} + \beta_{X_{t-2}} + \beta_{X_{t-3}} + \beta_{X_{t-4}} + \beta_{X_{t-5}} + \beta_{X_{t-6}} + \beta_{X_{t-7}} = 1.38.$$

- Transition model with lagged illness (response, order $= 2$: $Y_{it-k}, k = 1, 2$) and lagged maternal stress (order $= 3$: $X_{it-k}, k = 1, 2, 3$). GEE with independence working correlation matrix (Table 12.9, DHLZ).

```
> tran <- gee (illness ~ i1+i2+ s1 + s2 + s3
+               + employed + married +  mhlth +chlth
+               + race + education+ housize,
+              data = mscm, id = id,
+              family = binomial, na.action = na.exclude)
> summary(tran)

Coefficients:
                    Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept)       -2.9091220  1.0214378 -2.8480657   0.3963895 -7.3390500
i1                 2.3482833  0.1259368 18.6465240   0.1595907 14.7144149
i2                 0.3069020  0.1393614  2.2022019   0.1427095  2.1505358
s1                 0.2504857  0.1502302  1.6673457   0.1368713  1.8300817
s2                -0.1380240  0.1565727 -0.8815326   0.1555610 -0.8872661
s3                 0.4055328  0.1448058  2.8005283   0.1272986  3.1856814
employedemployed   0.1410396  0.1374382  1.0262037   0.1319936  1.0685332
marriedmarried     0.4248308  0.1237821  3.4320867   0.1174624  3.6167370
mhlthpoor         -0.3080916  0.5327726 -0.5782798   0.2057846 -1.4971561
mhlthfair         -0.3975853  0.5209759 -0.7631548   0.2182517 -1.8216828
mhlthgood         -0.3410129  0.5370923 -0.6349242   0.2460511 -1.3859434
mhlthvery.good    -0.2988771  0.5507248 -0.5426978   0.2441258 -1.2242747
chlthpoor          0.6900235  0.8372782  0.8241269   0.2170358  3.1793071
chlthfair          0.3218526  0.8295526  0.3879834   0.2037523  1.5796270
chlthgood          0.1670735  0.8249341  0.2025295   0.2555235  0.6538478
chlthvery.good    -0.1464070  0.8303261 -0.1763247   0.2491000 -0.5877439
racenonwhite       0.3490759  0.1246474  2.8005072   0.1351749  2.5824008
educationsomeHS    0.2210715  0.2297891  0.9620626   0.1845168  1.1981102
educationHS        0.1851319  0.2390116  0.7745726   0.2065900  0.8961316
educationsomeC     0.2676163  0.2731729  0.9796588   0.2223535  1.2035622
```

```
educationCG        0.7740110  0.4394452  1.7613366   0.3848820  2.0110343
housize>3         -0.4689530  0.1243052 -3.7725936   0.1256899 -3.7310314


Estimated Scale Parameter:  0.9800453
Number of Iterations:  1


Working Correlation
 ....
```

- g-computation: use the previous transition model for Y_{it}, identify a covariate value of interest $(\boldsymbol{Z}_i = \boldsymbol{z}^*)$, say (employed $= 0$, married $= 0$, maternal and child health $= 4$, race $= 0$, education $= 0$, and house size $= 0$), and generate Markov chains with stress controlled at 1 for all times, and controlled at 0 for all times.

We obtain

$$\hat{\mu}^{(1)} = 0.189$$
$$\hat{\mu}^{(0)} = 0.095$$

Hence, the causal log odds ratio is

$$\log\left(\frac{0.189/(1-0.189)}{0.095/(1-0.095)}\right) = 0.80$$

- MSM with IPTW: estimate $SW_i(t)$ based on models for the exposure process X_{it}, then use weighted GEE with independence correlation.

  1. For the denominator of the weight, use the logistic regression model to regression X_{it} on Y_{it}, Y_{it-1}, lagged stress $X_{it-1}, X_{it-2}, X_{it-3}$, and other baseline demographic covariates.

  2. For the numerator of the weight, use the same model except that Y's are excluded.

  3. Use the estimated weights in the working independence GEE.

<div align="center">

Table 12.11 in DHLZ

	log odds ratio[1]
Marginal association	1.38
Transition model	0.50
g-computation	0.80
MSM	0.85 (p=0.046)

</div>

[1]: for those employed=0, married=0, maternal and child health=4, race=0, education=0, and house size=0.

Further Reading

- Chapter 12 of DHLZ.

- Pepe MS and Anderson GA (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communication in Statistics - Simulation*, **23**(4):939–51.

- Robins JM, Greenland S, and Hu F-C (1999). Estimation of the causal effect of a time-varying esposure on the marginal mean of a repeated binary outcome. *JASA*, **94**:687-700.