Longitudinal Data Analysis II

PSYC 575

October 6, 2020 (updated: 17 January 2023)

Learning Objectives

- Specify models with alternative error **covariance structures**
- Describe the difference between analyzing trends vs. analyzing **dynamics** with longitudinal data
- Run analyses with **time-varying predictors** (i.e., level-1 predictors)
- Interpret and plot results

Covariance Structure

Longitudinal vs. Cross-Sectional Data

- School A: Student 1, 2, 3, . . .
  - Swapping the order is not a problem
- Person A: Observation 1, 2, 3 . . .
  - Swapping the order may be a problem
- Temporal ordering may mean that observations closer in time may be more strongly related
  - More likely when observations are a day apart vs. a year apart
- However, our previous models do not consider temporal dependence

Covariance Structure

- Covariances with respect to time
  - Covariance: how much two variables covary

```
> cov(curran_wide %>% select(read1:read4), use = "pair") %>% print(digits = 2)
      read1 read2 read3 read4
read1  0.86  0.67  0.57  0.48
read2  0.67  1.17  0.96  0.96
read3  0.57  0.96  1.35  1.08
read4  0.48  0.96  1.08  1.56
```

Is there a trend for the variances?

Complex Covariance Structure

- Serial Correlation
  - Correlation = covariance / (SD_1 * SD_2)

```
> cor(curran_wide %>% select(read1:read4), use = "pair") %>%
    print(digits = 2)
      read1 read2 read3 read4
read1  1.00  0.66  0.54  0.45
read2  0.66  1.00  0.78  0.76
read3  0.54  0.78  1.00  0.80
read4  0.45  0.76  0.80  1.00
```

Lag 2 correlation

Lag 1 correlation

How do the correlations look?

Complex Covariance Structure

- It is obvious that there are substantial covariances across time, and the variances seem to increase
  - Potential violations of (a) independent observations and (b) homogeneity of variance
- In MLM, the implied temporal covariance has the form

  ZGZ’ + **R**

  - **Z**: Design matrix for random effects
  - **G**: Covariance matrix of random effects (i.e., u_{0i}, u_{1i}, etc)
  - **R**: Covariance matrix of errors (i.e., e_{ti})

Covariance Structure in OLS

- OLS: Independence
  - Only **R**, with constant variance over time

1.09			
0.00	1.09		
0.00	0.00	1.09	
0.00	0.00	0.00	1.09

Covariance Structure in Random-Intercept MLM/Repeated Measures ANOVA

- **ZGZ'**
  - Same covariance for all time points

0.79			
0.79	0.79		
0.79	0.79	0.79	
0.79	0.79	0.79	0.79

+

- **R**
  - Independent and constant variance over time

0.41			
0	0.41		
0	0	0.41	
0	0	0	0.41

Covariance Structure in Random-Intercept MLM/Repeated Measures ANOVA

1.20			
0.79	1.20		
0.79	0.79	1.20	
0.79	0.79	0.79	1.20

Does this seem to describe the data well?

Covariance Structure With Random Slopes (Piecewise Growth)

- **ZGZ'**
  - Same covariance for all time points

0.60			
0.64	0.92		
0.62	1.00	1.13	
0.60	1.08	1.26	1.44

+

- **R**
  - Independent and constant variance over time

0.35			
0	0.35		
0	0	0.35	
0	0	0	0.35

Covariance Structure With Random Slopes

- Covariance

0.94			
0.64	1.26		
0.62	1.00	1.47	
0.60	1.08	1.26	1.79

- Correlation

1.00			
0.59	1.00		
0.53	0.73	1.00	
0.46	0.72	0.78	1.00

So far we have only looked at the **ZGZ'** part, which are due to person-specific intercepts and slopes

Autoregressive(1) Error Covariance Structure

- Decreasing correlation across time:
  - Lag 1 = ρ; Lag 2 = ρ^2
  - $-1 \le \rho \le 1$

$$\mathbf{R} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}$$

- Estimated $\rho = 0.04$

0.26			
0.01	0.26		
0.00	0.01	0.26	
0.00	0.00	0.01	0.26

R Output

```
> glmmTMB(read ~ piece(time0, node = 1) + (piece(time0, node = 1) | id) +
                  ar1(0 + factor(time) | id),
  dispformula = ~0, data = curran_long, REML = TRUE,
  control = glmmTMBControl(optimizer = optim, optArgs = list(method = "BFGS")
  ) %>%
  summary()
```

```
Random effects:

Conditional model:
 Groups Name           Std.Dev. Corr
 id     (Intercept)    0.7714
        phase1         0.4743    0.13
        phase2         0.2247   -0.12        0.96
 id.1   factor(time)1 0.5115   0.04 (ar1) 0.04 (ar1) 0.04 (ar1)
```

Error autocorrelation is small (0.04), after including the random slopes

Likelihood Ratio Test

```
anova(m_pw, m_pw_ar1)

Data: curran_long
Models:
m_pw: read ~ phase1 + phase2 + (phase1 + phase2 | id), zi=~0, disp=~1
m_pw_ar1: read ~ phase1 + phase2 + (phase1 + phase2 | id) + ar1(0 + factor(time) | , zi=~0, 
disp=~0
m_pw_ar1:     id), zi=~0, disp=~1
         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
m_pw     10 3229.7 3281.6 -1604.9   3209.7
m_pw_ar1 11 3231.6 3288.7 -1604.8   3209.6 0.0973      1      0.755
```

Autoregressive effect not significant

Remarks

- There are many other structures discussed in the longitudinal data analysis
  - E.g., AR(2), Toeplitz, etc
- The closer the time points are, the more likely that the errors have temporal correlations, even after including the random slopes
- In my experience, including an AR(1) structure does a reasonable job in many situations

Example

The Cognition, Health, and Aging Project

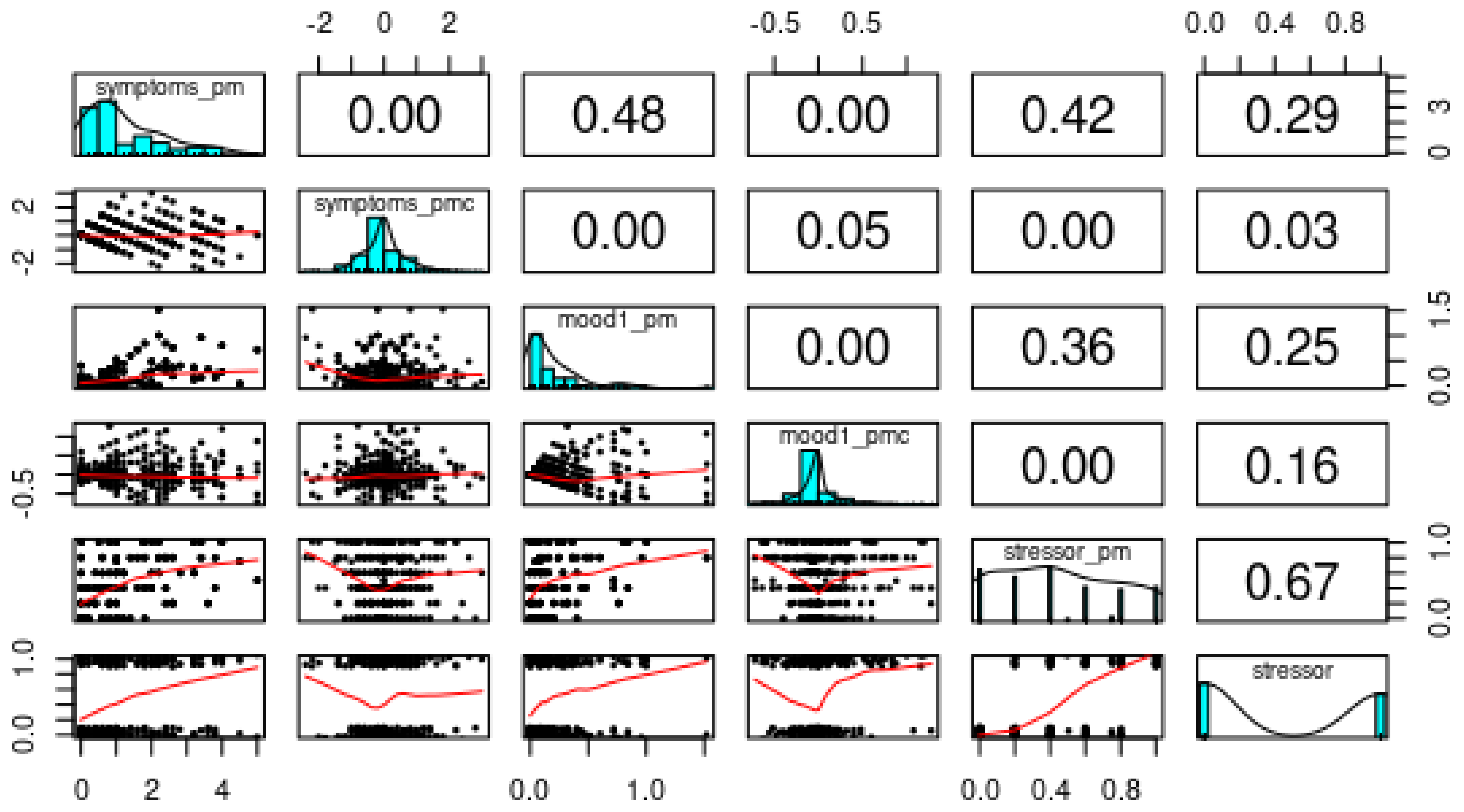
- The first wave of the CHAP
- Six observations over two weeks
  - Sessions 2-6
- baseage: $M = 80.13$ ($SD = 6.11$)

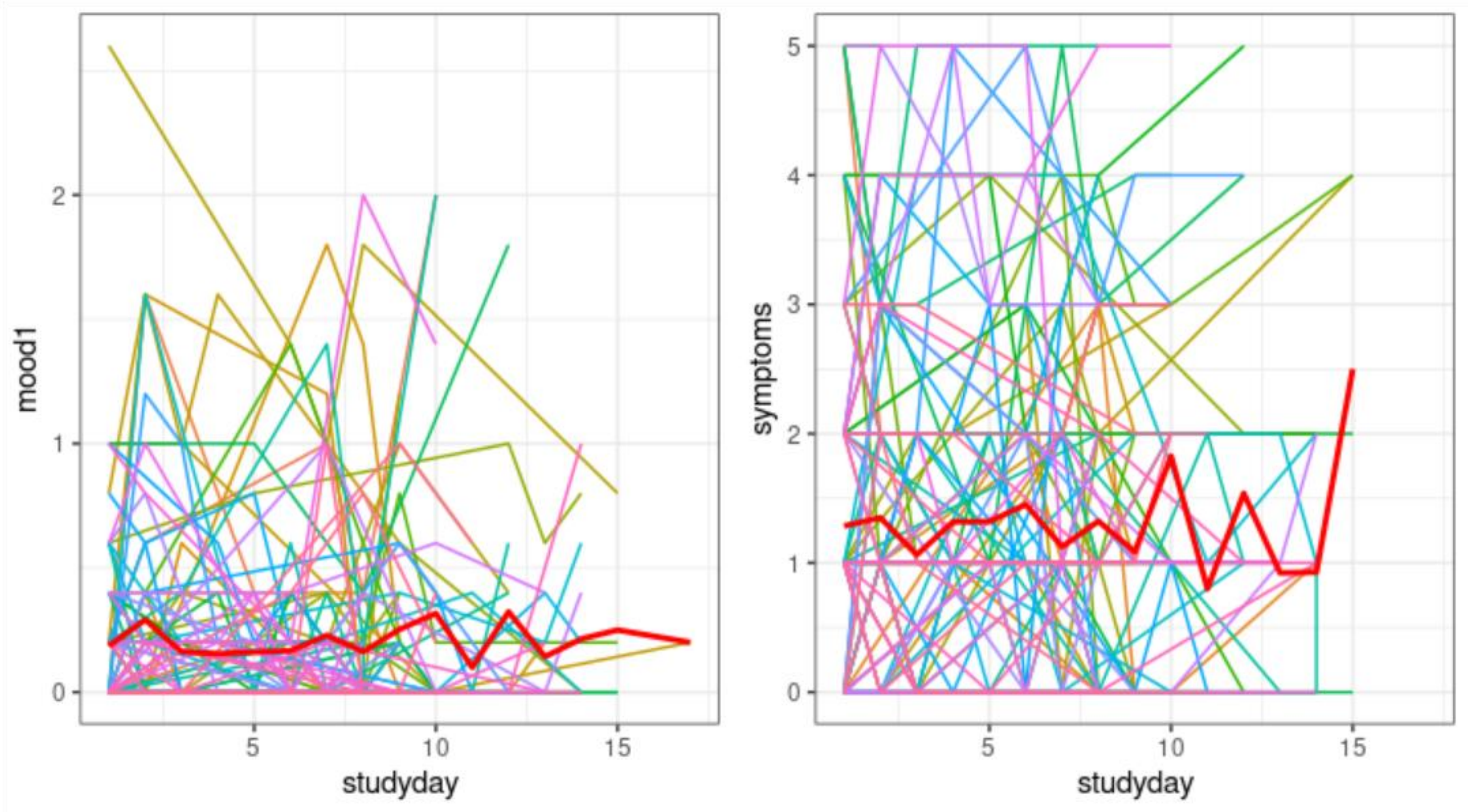
Time-Varying Covariates

- Variables at the within-person level that changes over time
- Need cluster-mean/person-mean centering
  - Between-person/within-person effects
- `Symptoms`: Number of physical symptoms in the past 24 hours
  - Max = 5
- `Mood`: Daily report negative mood (1 – 5)
  - `Mood1`: center at 1 (0 – 4)
- Stressor: Presence of a daily stressor (0 = stressor-free day; 1 = stressor day)

Decomposition of Effects

- Very important for some variables with longitudinal data
  - But not for the “time” variable
  - May not be meaningful for other measures of time (e.g., age)
- Trait: Person mean, time-invariant (in some sense)
- State: Deviation (fluctuation) from person mean, time-varying


symptoms_pm
symptoms_pmc
mood1_pm
mood1_pmc
stressor_pm
stressor
0.00
0.48
0.00
0.42
0.29
0.00
0.05
0.00
0.03
0.00
0.36
0.25
0.00
0.16
0.67



mood1
0
1
2
5
10
15
studyday
symptoms
0
1
2
3
4
5
5
10
15
studyday

Describing Fluctuations

- TIME may not be a predictor (unless a stable trend is found)
- The interest is in the momentary changes

Model 1

Model Equations

Level 1:

$$\text{symptoms}_{ti} = \beta_{0i} + \beta_{1i}\text{mood1\_pmc}_{ti} + e_{ti}$$

Level 2:

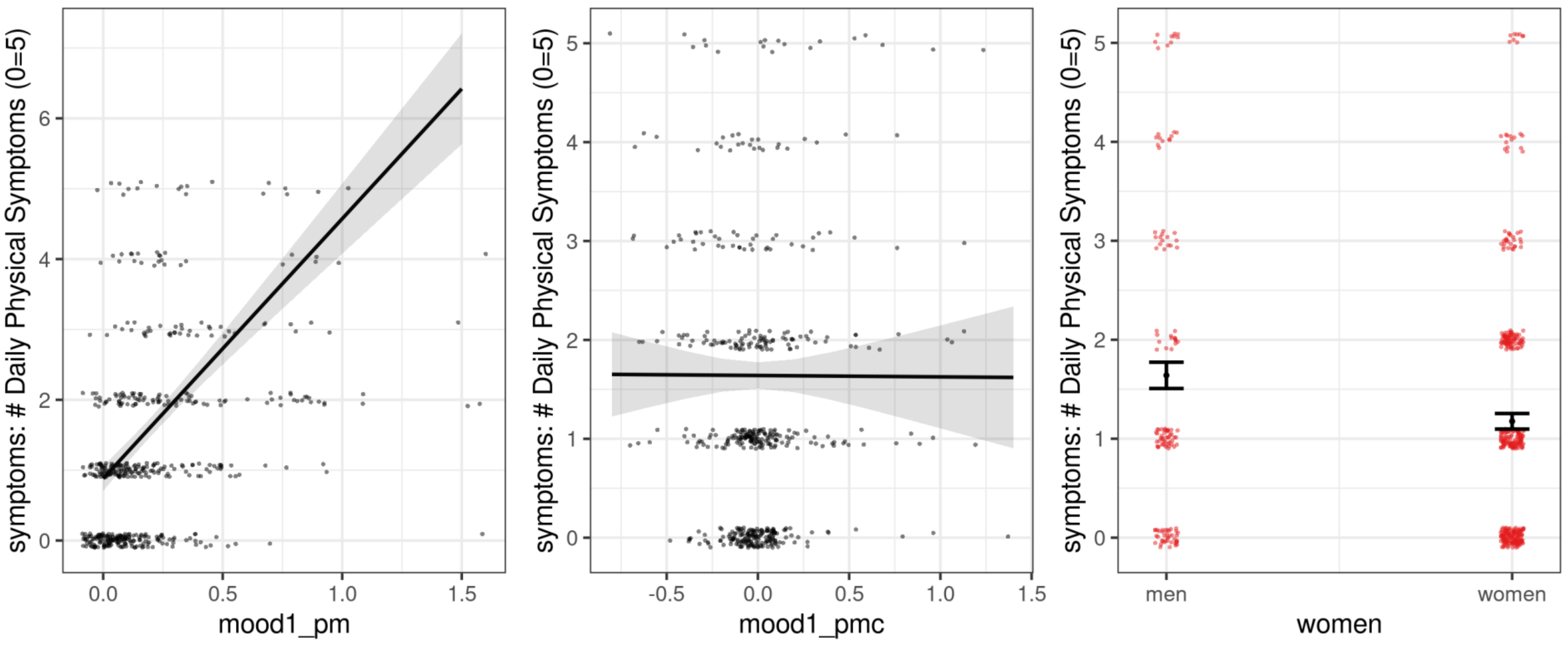
$$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{mood1\_pm}_i + \gamma_{02}\text{women}_i + \gamma_{03}\text{mood1\_pm}_i \times \text{women}_i + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{women}_i + u_{1i}$$

Fixed Effects (with `glmmTMB`)

```
Conditional model:
                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)             0.86403    0.24618   3.510 0.000449 ***
mood1_pm                3.86285    0.81368   4.747 2.06e-06 ***
mood1_pmc               0.00396    0.26835   0.015 0.988225    
womenwomen             -0.04167    0.28314  -0.147 0.883002    
mood1_pm:womenwomen    -2.14123    0.90529  -2.365 0.018018 *  
mood1_pmc:womenwomen    0.16552    0.30859   0.536 0.591699    
```

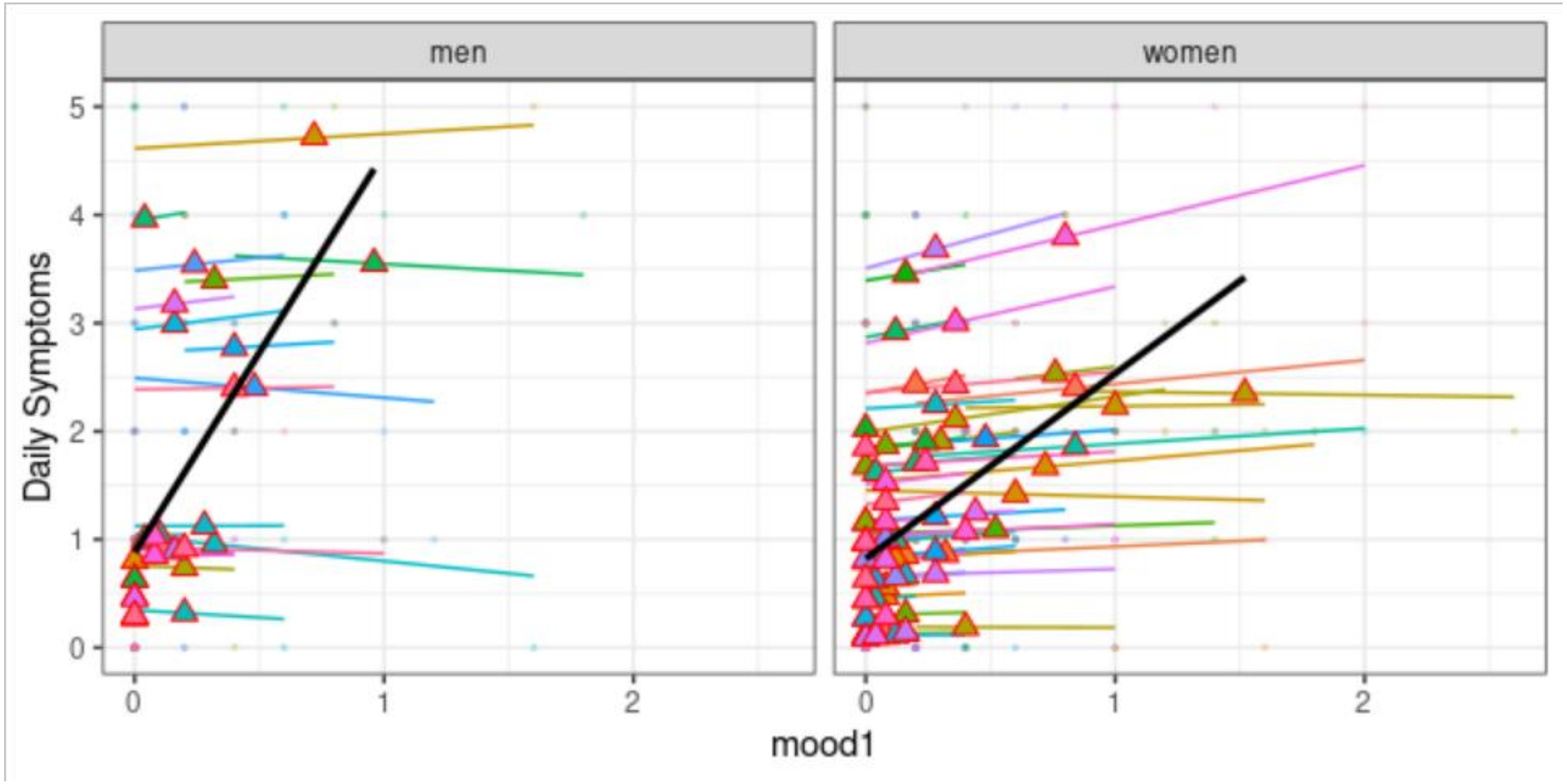
Note the between-person and the within-person effects are drastically different

plot_model(m1)



Interaction Plots

Between/Within Effects

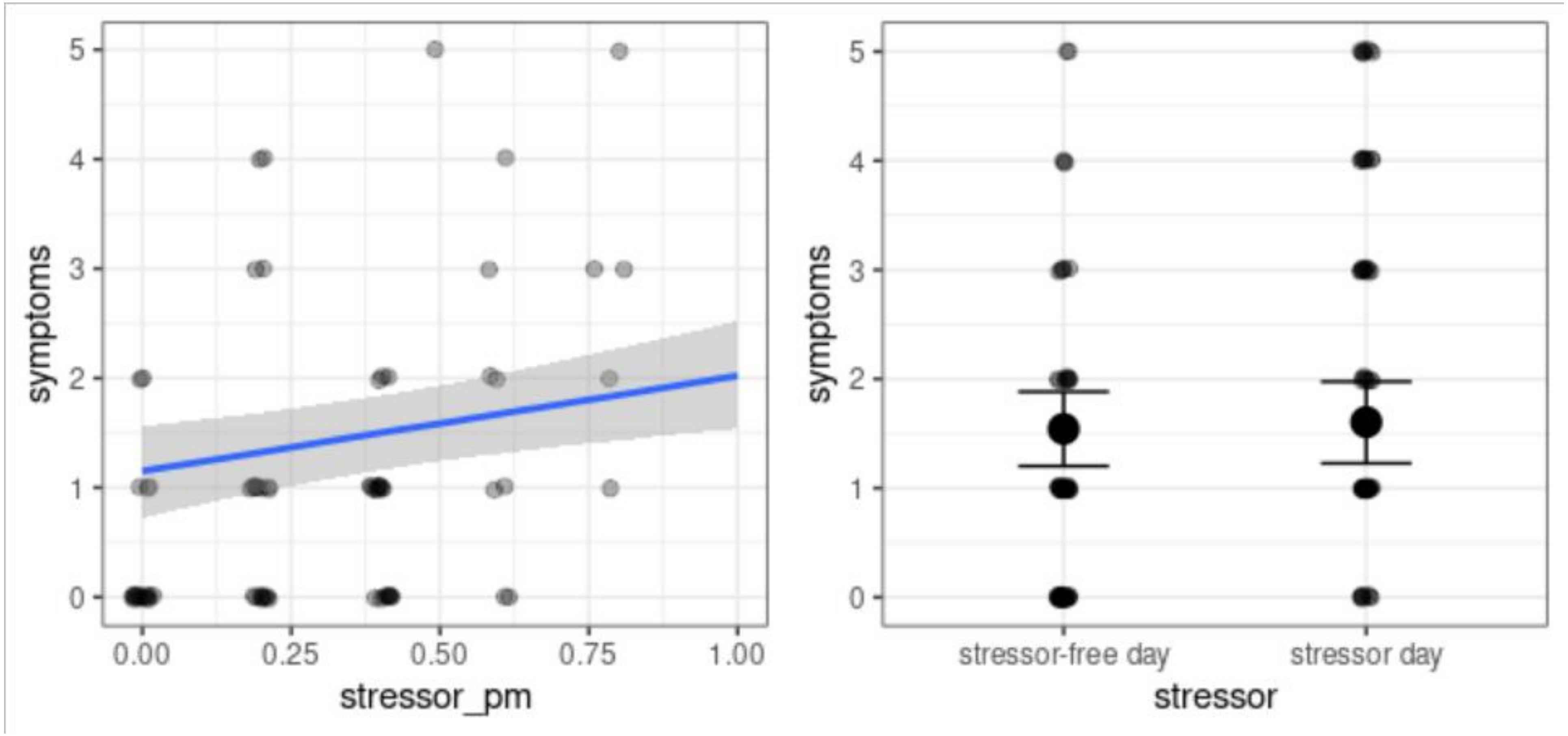


Model 2

Add `stressor` to the Equation

- A time-varying binary variable
- `stressor_pm` (person mean): Average stress level of a person (over the study period)
- However, the deviation from the person mean is harder to interpret
  - E.g., `stressor_pmc` = 0.8?
  - Methodologists do not agree on how to treat it, but for this example we'll keep the binary lv-1 variable
  - ➔ Contextual & within-person

Contextual and Within-Person Effects



Contextual Effect

```
Conditional model:
                     Estimate Std. Error z value Pr(>|z|)
...
stressor_pm            0.8487     0.3008    2.82   0.0048 **
stressorstressor day   0.0645     0.1005    0.64   0.5211
```

- On a stressor day (or a stressor-free day), a person who is one unit higher on average stress level reported on average 0.85 more symptoms, 95% CI [0.26, 1.44].

Topics Not Covered

- Comparable metric across time
  - Vertical scaling/Longitudinal measurement invariance
- Lag relationship/cross-lagged/autoregressive model
- Parallel-process model
- Missing data handling
- Multiple cohort design