

Inhaltsverzeichnis

1 Längsschnittanalysen und Veränderungsmessung II	1
1.1 Lernziele und Vorgehen	1
1.2 Vorbereitung	2
1.3 Datensatz	2
2 Quadratische Wachstumskurvenmodelle	3
2.1 Modellvergleiche	9
3 Aufnahme von Kovariaten	11

1 Längsschnittanalysen und Veränderungsmessung II

Im folgenden werden wir die im Skript zur Veränderungsmessung I betrachteten Wachstumskurvenmodelle im Rahmen der Mehrebenenanalyse erweitern. Zunächst werden wir eine Erweiterung auf nicht-lineares Wachstum betrachten. Im nächsten Schritt nehmen wir zeitinvariante und zeit-variiere Kovariaten in das Modell auf.

1.1 Lernziele und Vorgehen

Sie können Wachstumskurvenmodelle mit verschiedenen Formen der Veränderung und unter Berücksichtigung von Kovariaten spezifizieren und die Ergebnisse angemessen interpretieren.

Wir werden dafür wie folgt vorgehen:

Note

1. Betrachtung des Datensatzes (long-Format)
2. Wachstumskurvenmodelle im Rahmen der Mehrebenenanalyse mit lme4
 - a) Lineares Wachstum
 - b) Quadratisches Wachstum
 - c) Zeitstabile Kovariaten
 - d) Zeit-variiierende Kovariaten

1.2 Vorbereitung

Wir benötigen in dieser Sitzung die folgenden Pakete.

```
if (!require("lme4")) install.packages("lme4")
if (!require("ggplot2")) install.packages("ggplot2")
```

1.3 Datensatz

Wir nutzen den long-Format Datensatz zur Stimmungsregulation aus der letzten Sitzung (für detaillierte Informationen siehe Skript zur Veränderungsmessung I). In dieser Sitzung nutzen wir jedoch alle 12 vorhandenen Messzeitpunkte pro Person. Der Datensatz im long-Format (`affect.rda`) enthält die Stimmungsisems *gut* (valence), *wach* (energetic arousal), und *ruhig* (tense arousal), zwei dichotome Variablen (*pos* und *neg*) welche jeweils kodieren ob ein positives oder negative Ereignis stattgefunden hat seit der letzten Abfrage, sowie ein Maß für die Depressivität der Probandinnen (*BDI*; stabiler Wert, einmalig erfasst). Die Variable *ID* kodiert die Person, die Variable *MZP* nummeriert die Messungen pro Person durch, und die Variable *time* kodiert die seit der ersten Beobachtung vergangene Zeit kontinuierlich in Stunden.

```
setwd("C:/Users/me/myworkingdirectory")
load("affect.rda")
```

```
str(affect)
```

```
'data.frame':  797 obs. of  9 variables:
 $ ID    : Factor w/ 73 levels "182001","182002",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ time  : num  0.283 1.117 2.083 2.983 4.15 ...
 $ gut   : num  5 2 3.5 2 5 4.5 5.5 5 3 3 ...
 $ wach  : num  5.5 5 5.5 4.5 5.5 4.5 0 2.5 0 3.5 ...
 $ ruhig: num  2.5 2.5 2.5 2 3 3.5 5 4 4 1.5 ...
 $ pos   : num  0 0 0 0 1 0 1 0 1 1 ...
 $ neg   : num  1 0 1 0 0 0 0 1 0 0 ...
 $ BDI   : num  7 7 7 7 7 7 7 7 7 7 ...
 $ MZP   : int  1 2 3 4 5 6 7 8 9 10 ...
```



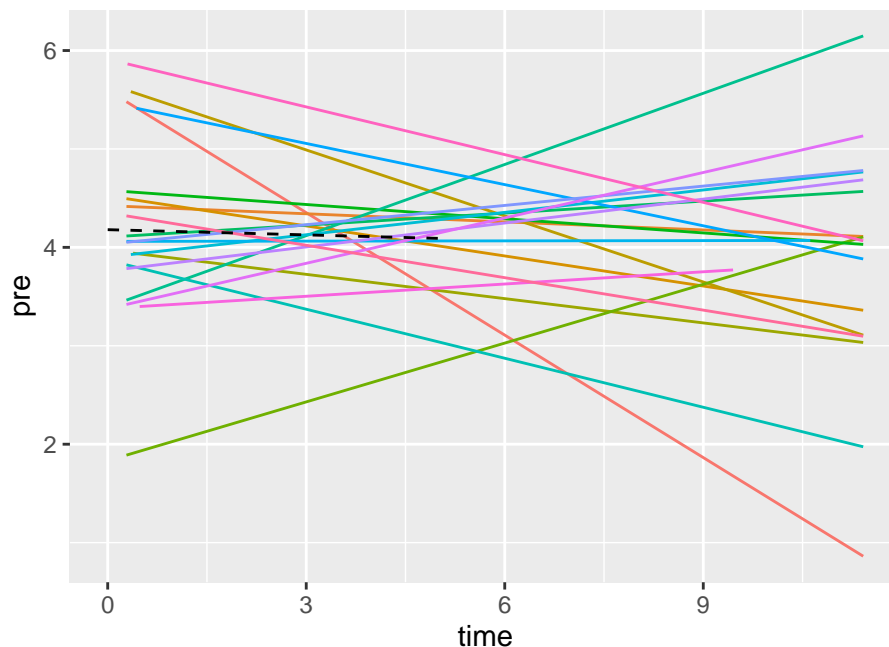
2 Quadratische Wachstumskurvenmodelle

Wir schätzen zunächst ein lineares Wachstumskurvenmodell mit der Variable Zeit (`time`) als Prädiktor. Lineare Wachstumskurvenmodelle oder auch sogenannte Latent-Growth-Curve Modelle nehmen in ihrer einfachsten Form an, dass der Wert einer Person über den Verlauf der Zeit kontinuierlich zu- oder abnimmt. Jede Person kann dabei eine individuelle Ausprägung (Betrag und Vorzeichen) des Anstiegs (bzw. der Abnahme) aufweisen.

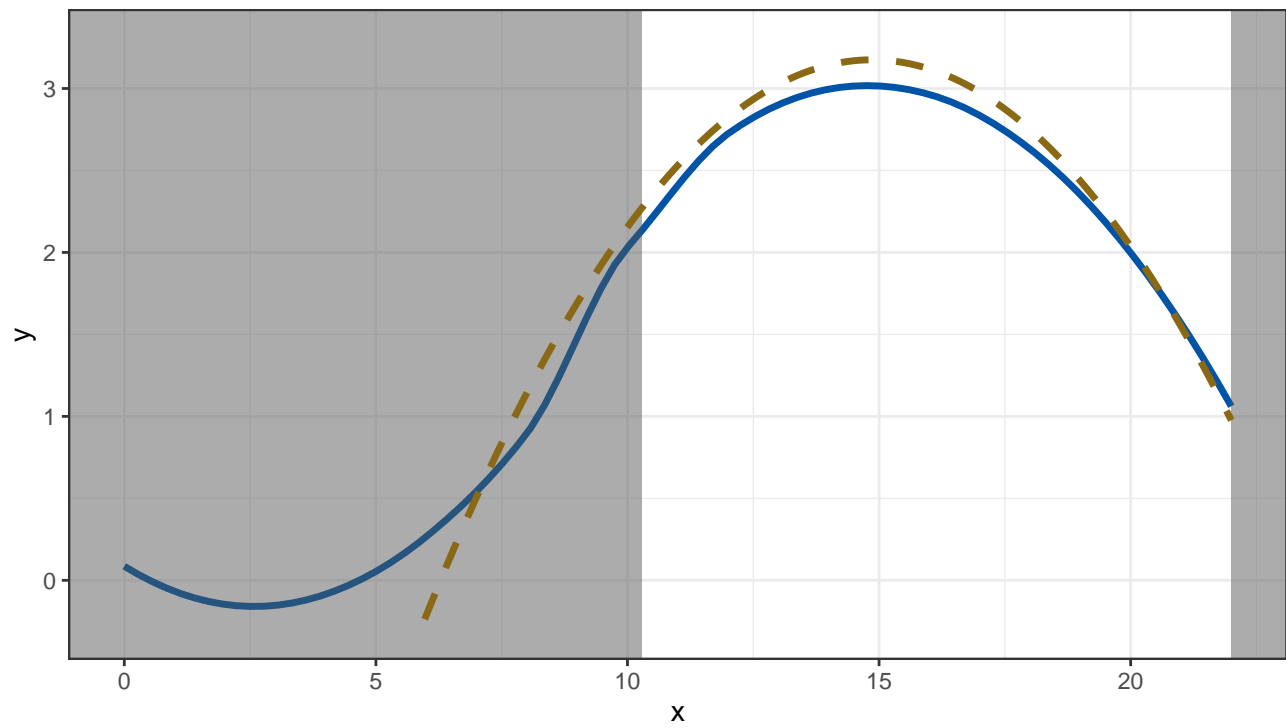
```
linear_RS <- lmer(wach ~ 1 + time + (1 + time | ID), REML = T, data = affect)
```

Wie bereits in der letzten Sitzung für die ersten 5 Messzeitpunkte gesehen, ist auch für die gesamten Daten eines Tages (alle 12 MZPe) ein Wachstumskurvenmodell mit individuell unterschiedlichen (random) Slopes für diese Daten angemessener als ein reines Variabilitätsmodell (Modellvergleich nicht gezeigt; siehe letzte Sitzung bzw. try it out!).

Wir plotten zur Veranschaulichung die Verläufe (einiger Personen) im random slopes Wachstumskurvenmodell (Code not shown):



Realistischer als der hier bisher modellierte lineare Verlauf über den Tag hinweg ist eine zirkadiane Verlaufsform, welche die Tageszeit berücksichtigt. Ein möglicher typischer Verlauf der Wachheit / Müdigkeit über den Tag hinweg könnte wie folgt aussehen (blaue Linie). Wenn man berücksichtigt, dass wir nur Daten während des Tages (nicht in der Nacht) erhoben haben, könnte aus theoretischer Perspektive ein quadratischer Verlauf approximativ ausreichend sein (gestrichelte Linie).



Um diesen Verlauf zu modellieren, erweitern wir das Modell um eine quadratische Komponente. Hierfür legen wir zunächst das Quadrat der Variable `time` als neue Variable an und nehmen diese dann als zusätzlichen Prädiktor ins Modell auf.

```
affect$time_quad <- affect$time^2
```

```
quad_RS_1 <- lmer(wach ~ 1 + time + time_quad + (1 + time | ID), REML = T, data = affect)
```

```
summary(quad_RS_1)$coef
```

	Estimate	Std. Error	t value
(Intercept)	3.81887493	0.182640129	20.909287
time	0.17422137	0.045239855	3.851060
time_quad	-0.01631269	0.003328546	-4.900846

```
summary(quad_RS_1)$var
```

Groups	Name	Std.Dev.	Corr
ID	(Intercept)	1.29740	
	time	0.17301	-0.784
Residual		1.00670	

```
confint(quad_RS_1,method = "boot", n = 1000)
```

Computing bootstrap confidence intervals ...

	2.5 %	97.5 %
.sig01	1.01832680	1.566995699
.sig02	-0.87343580	-0.648789581
.sig03	0.13794433	0.209086612
.sigma	0.95354279	1.064249670
(Intercept)	3.45494274	4.198536636
time	0.07994702	0.259991765
time_quad	-0.02292230	-0.009759218

Bedeutung der Parameter (siehe auch Unterlagen des Wintersemesters zu quadratischen Zusammenhängen):

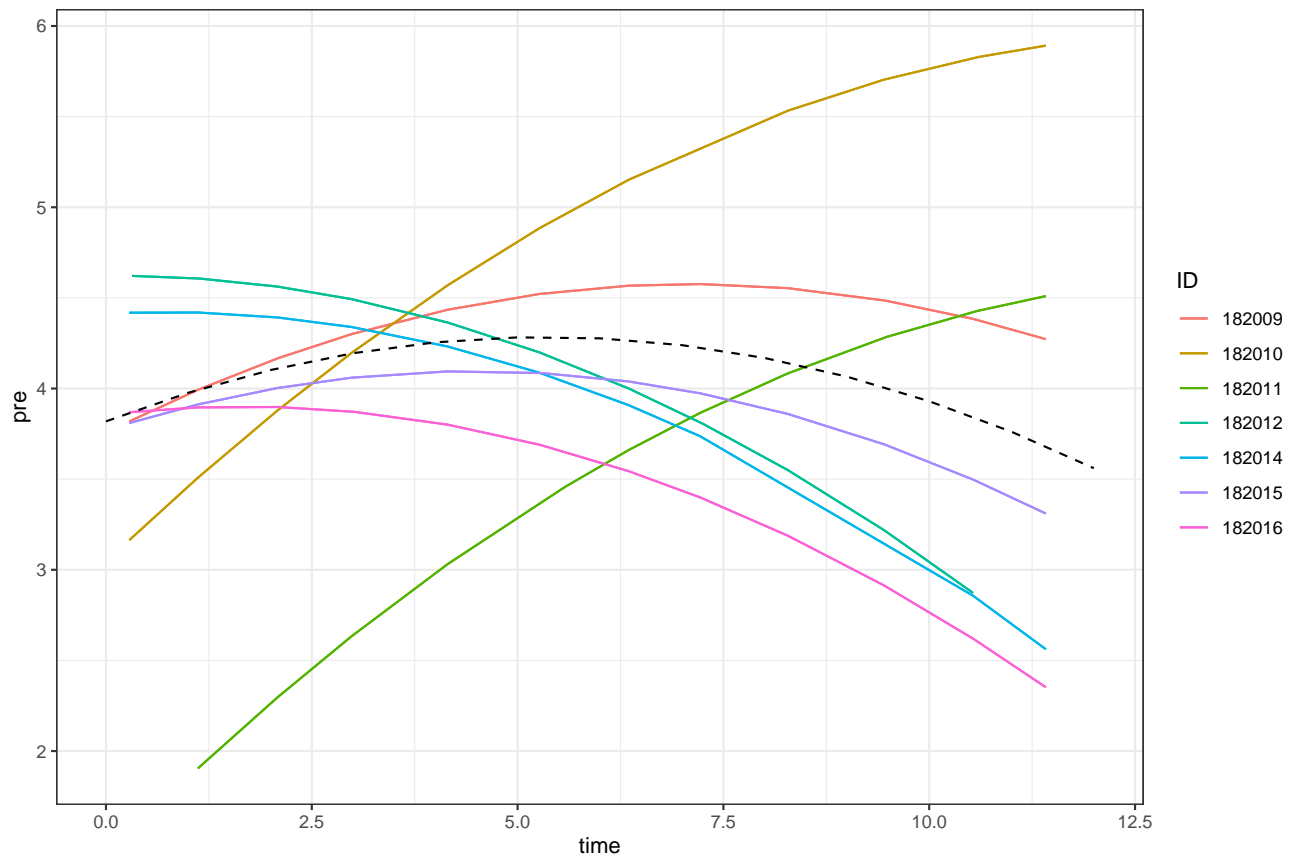
- (Intercept): (mittlerer) Erwartungswert zum Zeitpunkt 0 ($E[\text{wach} \mid \text{time} = 0]$) und Streuung der personenspezifischen Erwartungswerte über Personen
- time: Mittlere lineare Komponente des Wachstums über Personen hinweg sowie Streuung der personen-spezifischen linearen Komponenten
- time_quad: Quadratische Krümmung des Wachstums (fix über Personen)
- Corr: Korrelation der beiden personenspezifischen zufälligen Effekte (Intercepts und lineare Slopes)

Den fixed effects entsprechend können wir sehen, dass im Mittel über alle Personen hinweg die Wachheit zu Beginn des Tages ($\text{time}=0$) bei 3.82 liegt (auf der entsprechenden Skala). An den Vorzeichen und Werten der fixed effects der linearen und der quadratischen Komponente können wir ablesen, dass wir im Mittel über alle Personen hinweg erwarten, dass die Wachheit am morgen zunächst zunimmt und dann diese Zunahme aber kontinuierlich abflacht und schließlich zu einer Abnahme der Wachheit wird (siehe auch Grafik unten). Die Slopes des Prädiktors `time` variieren über Personen mit einer Standardabweichung von 0.173 mit einem 95% Bootstrap-KI=[0.138; 0.209] welches nicht die Null überdeckt. Es scheinen inter-individuelle Unterschiede im der linearen Komponente des Verlaufs der Wachheit über den Tag hinweg vorzuliegen. Die Wachheit mit der die Personen in den Tag starten (Intercept; $\text{time} = 0$) hängen zudem negativ mit ihrem linearen Wachstum über die Zeit zusammen ($r = -0.78$), das heißt, Personen, die wacher in den Tag starten zeigen tendenziell eher einen geringeren Anstieg bzw. einen Abfall der Wachheit über den Tag hinweg, während Personen die weniger wach in den Tag starten tendenziell einen höheren Anstieg der Wachheit über den Tag zeigen. Exemplarische Veranschaulichung der Verläufe einiger Personen (die schwarz gestrichelte Linie zeigt den mittleren Verlauf (beschrieben durch die fixed effects)):

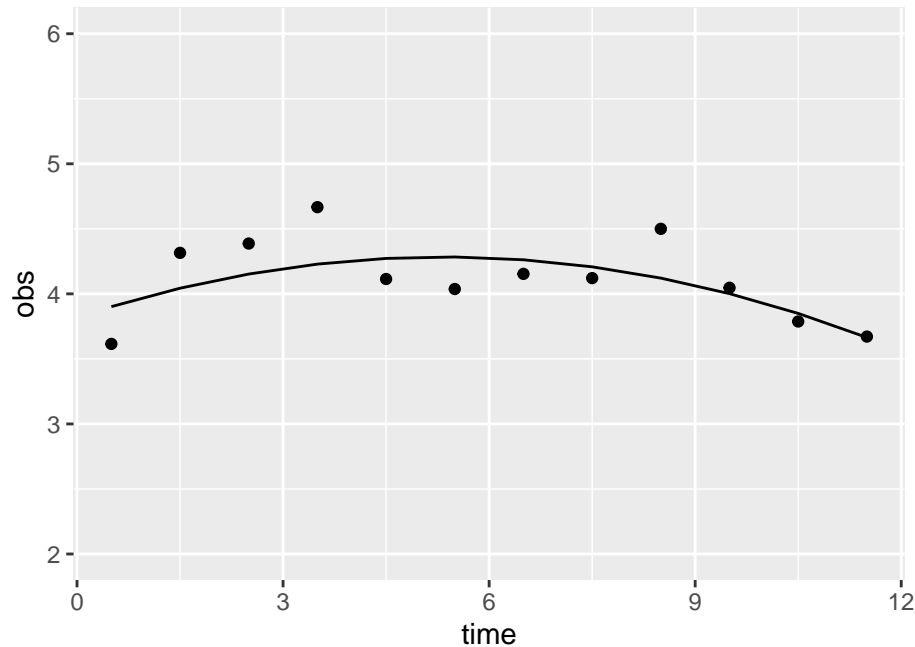
Wir berechnen die zeitpunktspezifischen Erwartungswerte (für ausgewählte Messzeitpunkte), die im quadratischen Modell erwartet werden und vergleichen diese mit den beobachteten Mittelwerten (für dieselben MZPe) (tabellarisch und gemeinsam Darstellung mit der vorhergesagten Verlaufskurve in einer Grafik).

```
time <- 0.5:11.5
fixed <- fixef(quad_RS_1)
pred <- fixed[1] + fixed[2]*time + fixed[3]*time^2
obs <- tapply(affect$wach, cut(affect$time, 0:12), mean, na.rm = T)
round(rbind(obs, pred), digits = 3)
```

	(0,1]	(1,2]	(2,3]	(3,4]	(4,5]	(5,6]	(6,7]	(7,8]	(8,9]	(9,10]	(10,11]
obs	3.615	4.315	4.387	4.667	4.114	4.037	4.154	4.121	4.500	4.047	3.788
pred	3.902	4.044	4.152	4.229	4.273	4.284	4.262	4.208	4.121	4.002	3.850
	(11,12]										
obs	3.672										
pred	3.665										



```
library(ggplot2)
tmp <- data.frame(time,obs,pred)
ggplot(tmp,aes(x=time,y=obs)) + geom_point() + geom_line(aes(y=pred)) + ylim(2,6)
```



Wir könnten uns nun fragen, ob auch personenspezifische Unterschiede in den quadratischen Komponenten des Wachstums vorliegen. Um dies zu überprüfen könnten wir einen random effect für den Prädiktor `time_quad` in das Modell aufnehmen:

```
quad_RS_1 <- lmer(wach ~ 1 + time + time_quad + (1 + time + time_quad | ID), REML=T, data=aff)
```

Warnung: Model failed to converge with max|grad| = 0.0917663 (tol = 0.002, component 1)

Warnung: Model is nearly unidentifiable: very large eigenvalue

- Rescale variables?

Die Warnungen die auftreten könnten auf Probleme hindeuten, die durch das Varianzenverhältnis von `wach` (2.005) zu `time_quad` (1856.017) entstehen. Ein erster Lösungsansatz besteht darin, die Zeitvariable zu reskalieren und die Modelle dann erneut anzupassen. Wir reskalieren die Variable `time` so, dass sie Zeit nicht mehr in Stunden, sondern zweistündig kodiert und wiederholen die Berechnung

von `time_quad` und Schätzung der Modelle entsprechend. Die Standardeinstellungen der Konvergenzkriterien (Toleranzen) der Optimizer in `lme4` sind zudem teilweise sehr streng, ggf. kann ein anderer Optimizer getestet werden:

```
affect$time <- affect$time/2
affect$time_quad <- affect$time^2
quad_rs1 <- lmer(wach ~ 1 + time + time_quad + (1 + time | ID), affect)
quad_rs2 <- lmer(wach ~ 1 + time + time_quad + (1 + time + time_quad | ID), affect,
                 control = lmerControl(optimizer="bobyqa"))
```

Die Warnung verschwindet. Zufällige quadratische Trends:

2.1 Modellvergleiche

Wir möchten nun Modellvergleiche heranziehen, um zu untersuchen, ob wir die quadratische Trend-Komponente und die zufälligen Effekte für diese Komponente im Modell brauchen, ein Modell ohne diese Komponenten also einen schlechteren Modell-Fit aufweist. Generell könnten wir unsere Modelle schrittweise aufbauen und gegeneinander testen, z.B. mit folgendem Vorgehen (Achtung, Modelle von oben nochmal geschätzt wegen anderer Skalierung der Zeit):

```
line <- lmer(wach ~ 1 + time + (1 | ID), affect)
quad <- lmer(wach ~ 1 + time + time_quad + (1 | ID), affect)
quad_RS_1 <- lmer(wach ~ 1 + time + time_quad + (1 + time | ID), affect)
quad_RS_2 <- lmer(wach ~ 1 + time + time_quad + (1 + time + time_quad | ID), affect)
```

Modellvergleich `line` vs. `quad` mit FIML estimation:

```
anova(line, quad, refit = TRUE)
```

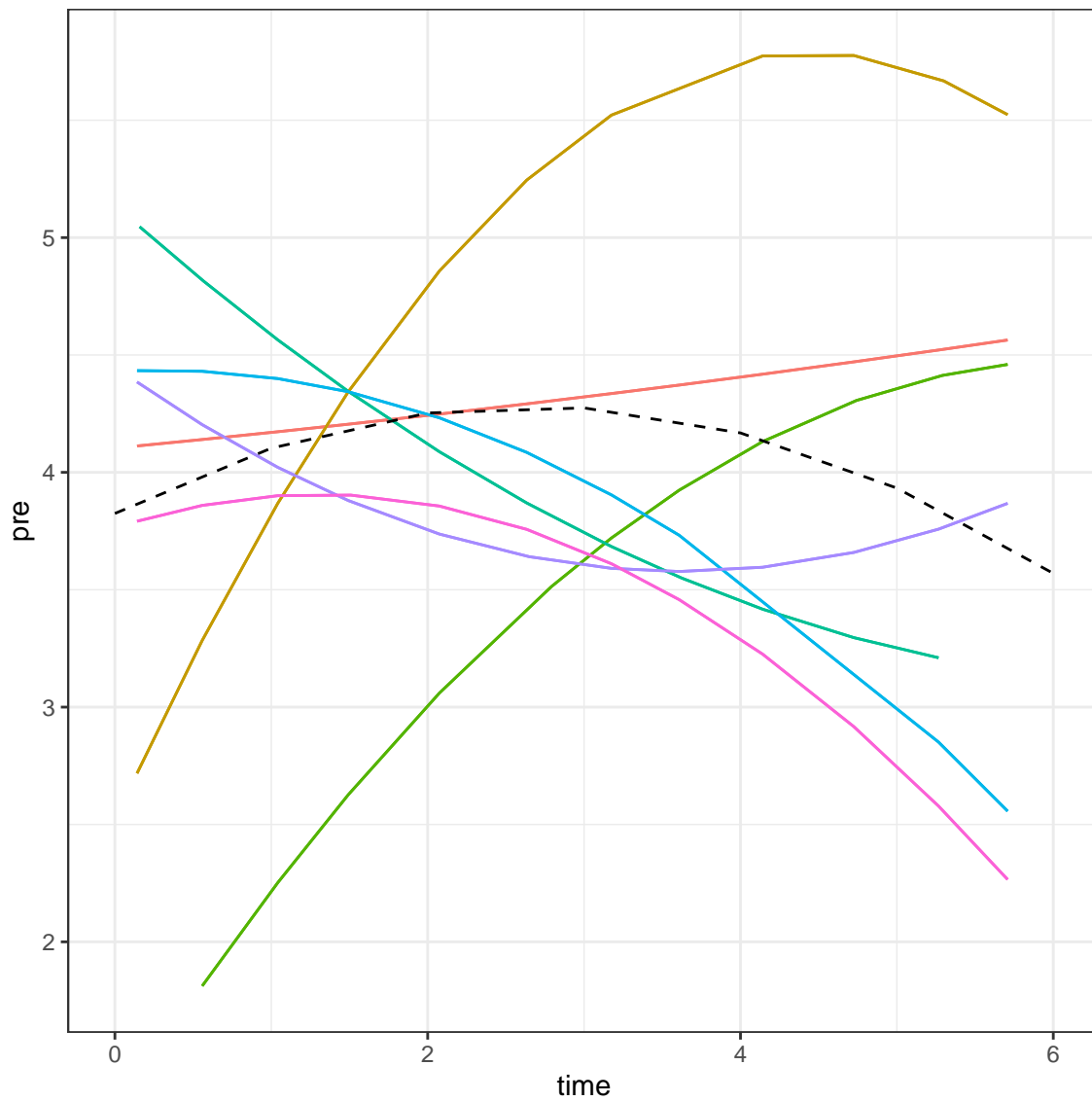
refitting model(s) with ML (instead of REML)

Data: affect

Models:

```
line: wach ~ 1 + time + (1 | ID)
quad: wach ~ 1 + time + time_quad + (1 | ID)
```

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
line	4	2678.1	2696.8	-1335.1	2670.1			
quad	5	2663.9	2687.3	-1327.0	2653.9	16.216	1	5.653e-05



Der quadratische Effekt der Zeit reduziert als zusätzlicher Prädiktor die Devianz bedeutsam, die Nullhypothese wird verworfen, $\chi^2(1) = 16,2; p < 0,001$.

Ist es notwendig, Unterschiede in den Regressionsgewichten der Zeiteffekte zuzulassen? Da es sich bei den zusätzlichen Parametern um zufällige Effekte handelt, soll die REML-Schätzung für den Vergleich beibehalten werden (`refit = FALSE`). Modellvergleich `quad` vs. `quad_RS_1` vs. `quad_RS_2` mit REML Schätzung:

```
anova(quad,quad_RS_1,quad_RS_2, refit = FALSE)
```

Data: affect

Models:

`quad: wach ~ 1 + time + time_quad + (1 | ID)`

`quad_RS_1: wach ~ 1 + time + time_quad + (1 + time | ID)`

`quad_RS_2: wach ~ 1 + time + time_quad + (1 + time + time_quad | ID)`

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
quad	5	2678.8	2702.2	-1334.4	2668.8			
quad_RS_1	7	2565.0	2597.7	-1275.5	2551.0	117.8572	2	< 2e-16
quad_RS_2	10	2562.6	2609.4	-1271.3	2542.6	8.3683	3	0.03898

Die Aufnahme der zufälligen linearen Regressionsgewichte reduziert die Devianz ($\chi^2(2) = 118; p < 0,001$). Die zusätzliche Aufnahme der zufälligen quadratischen Regressionsgewichte über den zufälligen linearen Effekte hinaus reduziert die Devianz noch weiter ($\chi^2(3) = 8,37; p = 0,039$).

(Wieder Achtung: Varianzen werden bei diesen Tests auf Randwerte gesetzt, aber der kritische χ^2 -Wert nach Snijders & Bosker (2012) für $\alpha = 0,05$ in einem Modell mit einem random slope (~ 5.14) bzw. zwei random slopes (~ 7.05) wird in beiden Fällen überschritten, und auch AIC und BIC sprechen für das `quad_RS_1` Modell im Vergleich zum `quad` Modell (sind sich jedoch uneinig bzgl. `quad_RS_2`).

3 Aufnahme von Kovariaten



Im nächsten Schritt möchten wir Kovariaten in das Modell aufnehmen, die potentiell inter-individuelle Unterschiede in den Verläufen über die Zeit erklären können. Wir betrachten zunächst eine zeit-variierende Kovariate. Als solche liegt uns im Datensatz unter anderem die Variable `pos` vor, welche kodiert, ob die Person zu diesem Zeitpunkt ein positives Ereignis erlebt hat. Es handelt sich also um einen dichotomen Prädiktor, welcher die Werte 0 (kein pos. Ereignis) und 1 (pos. Ereignis hat stattgefunden) annehmen kann (Dummy Variable). Wir können diese direkt als Level-1 Prädiktor in das (quadratische) Wachstumskurvenmodell integrieren:

```
pred_L1 <- lmer(wach ~ 1 + time + time_quad + pos + (1 + time + time_quad | ID), affect)
summary(pred_L1)$coef
```

	Estimate	Std. Error	t value
(Intercept)	3.77198766	0.18315610	20.594388
time	0.33276475	0.10753352	3.094521
time_quad	-0.06283112	0.01635682	-3.841279
pos	0.27755143	0.09829427	2.823679

Im Mittel über alle Personen hinweg erwarten wir, dass der Eintritt eines positiven Ereignisses mit einer erhöhten Wachheit zu diesem Zeitpunkt einhergeht (Differenz in erwarteter Wachheit wenn pos. Ereignis vorliegt vs. nicht vorliegt = 0.278). In der grafischen Veranschaulichung des Modells kann man erkennen, dass der dichotome zeit-variable Prädiktor “Sprünge” in den vorhergesagten Trajektorien bewirkt.

Wir möchten nun überprüfen, ob Personen mit hohen vs. niedrigen Ausprägungen der Depressivität (zeitstabile Kovariate BDI), sich in ihren mittleren Werten oder ihren Verläufen der Wachheit / Müdigkeit über den Tag hinweg voneinander unterscheiden.

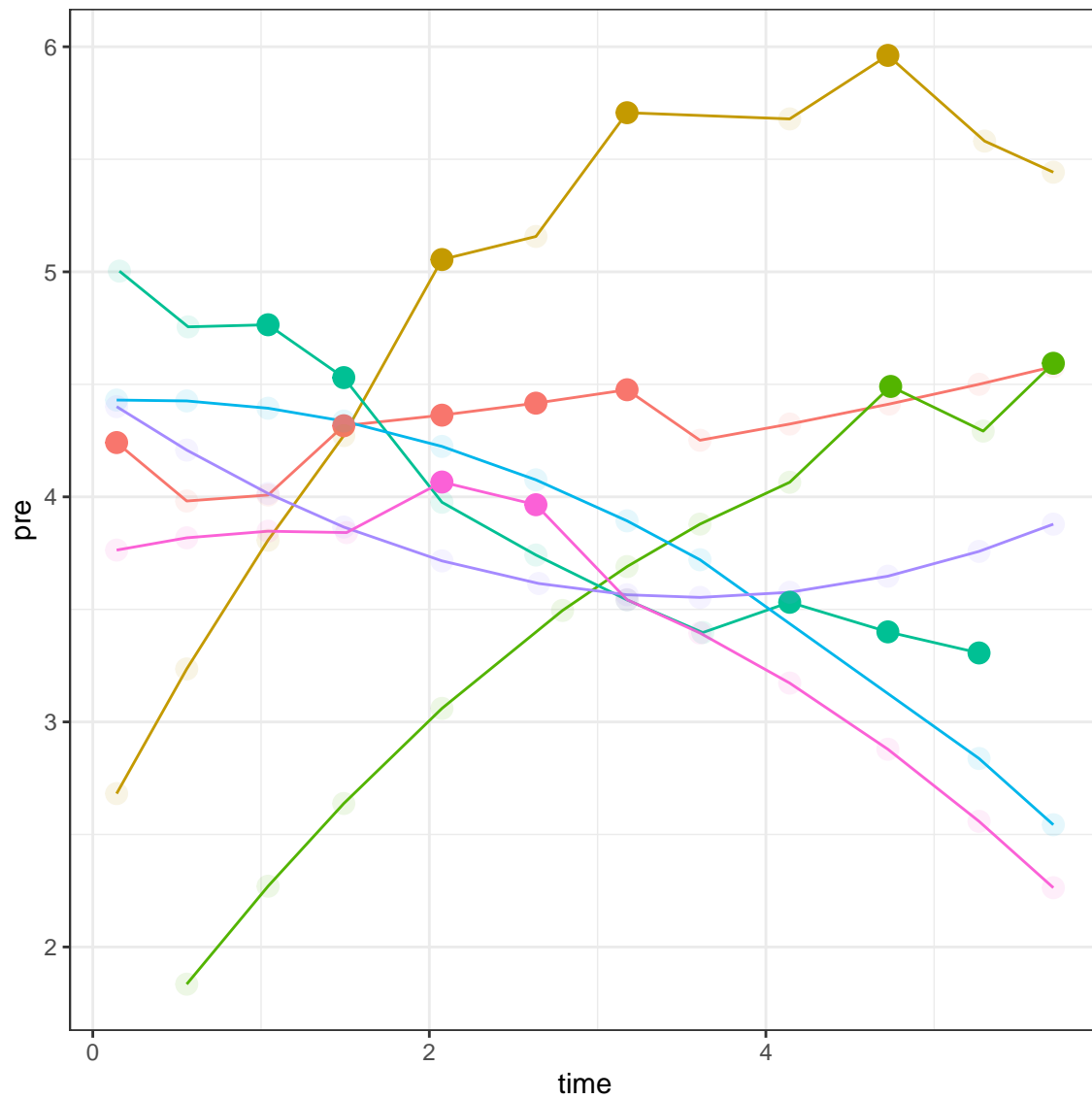
```
pred_L2 <- lmer(wach ~ 1 + time + time_quad + BDI + (1 + time + time_quad | ID), affect,
               control = lmerControl(optimizer="bobyqa"))
summary(pred_L2)$coef
```

	Estimate	Std. Error	t value
(Intercept)	3.96873301	0.19799059	20.045059
time	0.34330213	0.10688746	3.211809
time_quad	-0.06424264	0.01619789	-3.966112
BDI	-0.07298852	0.03721478	-1.961278

Der BDI (Summenscore) scheint kein bedeutsamer Prädiktor für Unterschiede im Ausgangsniveau der Wachheit zu sein. Ein negativer Effekt hätte hier z.B. bedeutet, dass, umso größer der BDI ist, desto geringer im Mittel die Wachheit einer Person (bzw. größer die Müdigkeit; im LST-Kontext: Zusammenhang auf Trait-Ebene).

Modell mit Cross-Level Interaktion:

```
affect$BDI <- scale(affect$BDI, scale = FALSE)
cross <- lmer(wach ~ 1 + time*BDI + time_quad + (1 + time + time_quad | ID), affect,
             control = lmerControl(optimizer="bobyqa"))
```



```
summary(cross)$coef
```

	Estimate	Std. Error	t value
(Intercept)	3.8234303469	0.18492157	20.67595682
time	0.3434546571	0.10707560	3.20759024
BDI	-0.0703463335	0.06341293	-1.10933745
time_quad	-0.0642574308	0.01619851	-3.96687283
time:BDI	-0.0009055944	0.01739647	-0.05205622

```
confint(cross,method = "boot", n = 1000)
```

Computing bootstrap confidence intervals ...

	2.5 %	97.5 %
.sig01	1.03501465	1.61331693
.sig02	-0.82550351	-0.33552778
.sig03	-0.36814918	0.59326571
.sig04	0.36662183	0.81575695
.sig05	-0.92679043	-0.48595429
.sig06	0.03375066	0.11542647
.sigma	0.92427511	1.03271420
(Intercept)	3.43080052	4.17259093
time	0.13994728	0.56265823
BDI	-0.19442619	0.05264723
time_quad	-0.09533017	-0.03434531
time:BDI	-0.03447421	0.03493696

Interpretation der Koeffizienten:

- time: Lineare Wachstumskomponente bei mittlerem BDI (zentriert)
- BDI: Einfluss des BDI auf die wache vs. müde Stimmung zum Referenzzeitpunkt
- time*BDI: Veränderung der linearen Wachstumskomponente pro Einheit in BDI
 - Der BDI scheint keinen Einfluss auf den linearen Verlauf der Wachheit über den Tag zu haben
 - ein negativer Effekt hätte hier z.B. bedeutet, dass, umso größer der BDI ist, desto schwächer der lineare Anstieg der Wachheit über den Tag ist