

Fakulta sociálních studií
Masarykova univerzita
Katedra psychologie

Stanislav Ježek

Modelování heterogenity v růstových modelech

Habilitační práce

Brno 2018

Rád bych zde poděkoval všem, kteří se mnou museli být tak trpěliví.

OBSAH

Od modelu růstových křivek k modelu růstových směsí.....	5
I. Modely heterogenity růstových křivek.....	9
Analýza latentních růstových křivek a model růstových křivek.....	9
Analýza latentních tříd růstových křivek.....	10
Růstový mixture model.....	11
Růstové parametry a jejich zobrazování.....	13
Růstové modely vyšších řádů	14
Růstový model faktorů	15
Model faktorů křivek.....	17
Další příbuzné modely	18
II. Základní postup analýz hledající latentní třídy.....	20
1. Specifikace nepodmíněného modelu latentních růstových křivek.....	20
Specifikace růstové (strukturní) části modelu	21
Specifikace reziduální struktury (stochastická část modelu).....	25
2. Provedení analýzy latentních tříd růstových křivek	26
3. Specifikace růstového mixture modelu	27
4. Řešení problémů s odhadem parametrů modelů.....	28
5. Volba optimálního nepodmíněného modelu s třídami.....	29
Informační kritéria	30
Likelihood ratio test.....	31
Ukazatele založené na úspěšnosti klasifikace.....	33
Shrnutí.....	36
6. Zařazení kovariátů do růstového modelu.....	37
Jednokrokový postup.....	38
Tříkrokový přístup	40
BCH přístup.....	42
LTB přístup.....	43
Shrnutí.....	43
III. Analýza heterogenity vývojových křivek rizikového chování v adolescenci.....	46
Krok 1 - Specifikace nepodmíněného modelu růstových křivek	48
Krok 2 - Modely latentních růstových tříd na latent-base modelu	51
LCGA modely se stejnými růstovými bázemi napříč třídami.....	51
LCGA Modely s různými růstovými bázemi napříč třídami	55
Krok 3 - Růstové mixture modely založené na latent-base modelu	58
GMM se stejnými růstovými bázemi napříč třídami.....	58
GMM s různými růstovými bázemi napříč třídami.....	61
Krok 4 - Volba optimálního modelu.....	65

Krok 5 - Zařazení kovariátů do GMM modelů.....	66
Latentní třídy vývoju užívání návykových látek	69
IV. Analýza heterogenity vývojových křivek prožívané autonomie.....	71
Model latentních růstových křivek	77
Příprava dat	77
LGC modely autonomie.....	79
Modely latentních růstových tříd autonomie	85
LCGA lineárního modelu s homoskedastickým rezidui	85
LCGA modely s heteroskedastickými rezidui.....	89
LCGA kvadratické modely s heteroskedastickými rezidui napříč třídami.....	91
LCGA kvadratické modely s autokorelační reziduální strukturou	93
Růstové mixture modely autonomie.....	96
GMM lineárního růstového modelu autonomie s homoskedastickými rezidui.....	97
Růstový model druhého řádu - Curve-of-factors model.....	102
Příprava dat	104
Longitudinální invariance	105
Růstový model latentní autonomie (CFM).....	108
Model latentních růstových tříd CFM	111
Model růstových směsí CFM - SO-GMM	117
Shrnutí modelů autonomie.....	126
Závěry	128
Reziduální struktura jako žádoucí ohnisko pozornosti	129
Hybridní specifikace latentních tříd.....	130
Vypořádání se s chybějícími daty	131
Limity růstových mixture modelů	132
Technické a praktické zkušenosti	132
Literatura.....	134
Příloha 1 - Specifikace prezentovaných modelů v Mplus	141
Rizikové chování.....	141
Prožívaná autonomie.....	148
Příloha 2 - R skripty pro zpracování výstupů z Mplus a tvorbu grafů	161
Skripty pro zpracování analýz rizikového chování	161
Skripty pro zpracování analýz Autonomie pomocí SO-GMM na CFM	166
Šablona pro generování Mplus syntaxu pro sadu GMM modelů	170

Od modelu růstových křivek k modelu růstových směsí

Růstové modely jsou jedním ze základních způsobů analýzy vývojových dat. Popisují změny psychologické charakteristiky v čase a umožňují modelovat vliv relevantních proměnných na podobu růstu. Vedle toho však umožňují kvantifikovat variabilitu (heterogenitu) růstu mezi jednotlivci. Ta je mnohdy velmi vysoká. Jednotlivci se mohou výrazně lišit tempem růstu, střední či počáteční úrovní sledované charakteristiky, tvarem růstové trajektorie, nebo všemi těmito charakteristikami najednou. Čím složitější růstový model, tím více je parametrů, v nichž se jednotlivci mohou mezi sebou lišit. Tyto individuální rozdíly se v přístupu zaměřeném na proměnné (*variable-centered approach*) snažíme vysvětlit či predikovat pomocí kovariátů, které lze obvykle přímo zařadit do modelu. V *person-centered* přístupu (Bergman, Magnusson, Khouri, Magnusson, & Khouri, 2003) využíváme různé klasifikační techniky, jejichž cílem je popsat heterogenitu/variabilitu jako výsledek existence více skupin (tříd, klastrů, subpopulací), uvnitř kterých bychom mohli považovat vývoj za homogenní (tj. pro všechny členy přibližně stejný, lišící se pouze náhodnou chybou). Z klasifikační technik se k tomu účelu dnes více a více používají různé podoby modelování latentních tříd, u kterých pak lze v následných analýzách hledat kovariáty pravděpodobnostního členství v těchto třídách a další vztahy. Stále se také používají různé podoby shlukové analýzy, které nepracují s představou latentních tříd. Nemohu na tomto místě nezmínit habilitační práci Lídy Osecké o aplikaci shlukové analýzy v psychologii osobnosti, která před mnoha lety můj zájem o klasifikaci podnítila (Osecká, 1999). Model růstových směsí je jednou z možností, jak *variable-centered* a *person-centered* přístup skloubit do jednoho modelu (Muthén & Muthén, 2000).

Jedním ze základních cílů longitudinálního výzkumu v psychologii je popis vývoje člověka v různě dlouhých obdobích života. Oproti průřezovým výzkumům, které umožňují identifikovat změnu či vývoj populačních parametrů zkoumaných charakteristik v čase (tedy srovnávat kohorty), umožňují longitudinální designy zaznamenat změnu či vývoj sledovaných charakteristik v čase u jednotlivců. Produktem longitudinálního výzkumu tak jsou časové řady – vývojové trajektorie – jednotlivých účastníků výzkumu. Vzhledem k obvyklým charakteristikám longitudinálního výzkumu v psychologii jde ve smyslu počtu měření obvykle o velmi krátké časové řady; výjimkou nejsou ani pouhá dvě měření. Tato měření jsou v psychologii zatížena nezanedbatelnou chybou. Velmi často jsou předmětem zájmu charakteristiky, které v použitých měřících či ontologicky nabývají jen několika málo hodnot. Vzhledem k uvedeným charakteristikám, a také vzhledem k velkým individuálním rozdílům mezi lidmi v řadě charakteristik, jimiž se psychologie zabývá, jsou individuální vývojové trajektorie obvykle velmi rozmanité (viz např. růstové křivky autonomie v kapitole 3). Příznačně se společně zobrazení více či všech individuálních vývojových trajektorií někdy nazývá špagetový graf (*spaghetti plot*, Swihart et al., 2010), seriózněji empirický graf růstu (*empirical growth plot*, Singer & Willett, 2003). Naplnění cílů longitudinálního výzkumu tak znamená teoreticky opodstatněné hledání pravidelností, vzorců v získaném souboru (a v konečném důsledku populaci) individuálních vývojových trajektorií.

Etablovaným analytickým nástrojem k hledání pravidelností a souvislostí v souboru individuálních vývojových trajektorií jsou různé podoby analýzy růstových křivek. Jde o obecný přístup, který je realizován buď v rámci víceúrovňového (generalizovaného) lineárního modelu, a pak se nazývá *growth curve modelling* (GCM, česky lze modelování růstových křivek, Singer & Willett, 2003), nebo v rámci strukturního modelování, a pak se nazývá *latent growth curve analysis* (LGCA, česky analýza latentních růstových křivek, Muthén,

2003). I když se tyto dva přístupy v řadě konkrétností a možností liší, snaží se v základu odpovědět na stejné základní otázky. Tou první otázkou je, jaké jsou vlastně individuální vývojové trajektorie, zda lineární, kvadratické či vyššího řádu, nebo zda jsou v nich diskontinuity, a otázka, zda se jednotlivci mezi sebou liší pouze v parametrech trajektorie téže formy (např. vývoj všech je lineární, liší se jen tempem růstu a počátečním stavem), nebo zda mají různí jednotlivci vývojové trajektorie lišící se i svou formou (někdo lineární, jiní kvadratickou). Ptáme se tedy po míře a konkrétní podobě *heterogenity* individuálních vývojových trajektorií. Pokud se na základě teorie a/nebo dat domníváme, že lze všechny individuální vývojové trajektorie popsat jako spojitě větší či menší odchylky od populační průměrné růstové křivky, dostáváme jednoduchý *nepodmíněný model růstových křivek* (unconditional growth curve model). Ten předpokládá, že parametry růstové křivky jsou náhodné a normálně rozložené. Jednoduchý nepodmíněný lineární model růstových křivek lze v rámci víceúrovňového lineárního modelu formálně zapsat následovně:

$$Y_{ij} = \pi_{0i} + \pi_{1i}T_{ij} + \varepsilon_{ij}$$

$$\pi_{0i} = \gamma_{00} + \xi_{0i}$$

$$\pi_{1i} = \gamma_{10} + \xi_{1i},$$

kde \hat{Y}_{ij} je odhadovaná hodnota modelované proměnné v čase T_j jednotlivce i . Individuální parametry lineárního růstu π_{0i} (průsečík) a π_{1i} (směrnice, tempo růstu) jsou v modelu reprezentovány tak, že mají normální rozložení s průměry γ_{00} a γ_{10} a rozptyly σ_0^2 a σ_1^2 . V souladu se zvyklostmi regresního modelu předpokládáme také normální rozložení chyb odhadu ε_{ij} . Formálně to můžeme zapsat:

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right),$$

z čehož mimo jiné plyne, že průsečík a tempo růstu jako náhodné proměnné nemusí být nezávislé a v rámci modelu se standardně odhaduje i jejich kovariance σ_{01} . Kdybychom chtěli specifikovat kvadratický model, přibyl by kvadratický člen $\pi_{2i} T^2$ a i zde by byl regresní koeficient π_{2i} náhodným parametrem s normální rozložením s průměrem γ_{20} a rozptylem σ_2^2 .

V jednoduchém nepodmíněném modelu růstových křivek jsou tak individuální rozdíly mezi vývojovými trajektoriemi zachyceny výhradně prostřednictvím náhodných regresních koeficientů. Rozdíly mezi lidmi tak mohou být v počátečním stavu a v rychlosti růstu a tyto rozdíly mají normální rozložení. Platí tedy, že nejvíce lidí má vývojovou trajektorii blízkou té průměrné, a čím více je trajektorie vzdálená té průměrné, tím je méně pravděpodobná, resp. četná v populaci.

Reálně může nastat řada situací, kdy podmínky, které popisuje nepodmíněný model růstových křivek, nejsou splněny. Jednou z nich je situace, kdy vývojové trajektorie, resp. jejich parametry či sledované manifestní proměnné ve skutečnosti nejsou normálně rozložené. Další běžně nastávající situací je, že v populaci existují skupiny, které se svými průměrnými vývojovými trajektoriemi výrazně liší. Pokud tyto skupiny z teorie známe, můžeme členství ve skupině zahrnout do modelu a umožnit různým skupinám mít odlišné průměrné růstové trajektorie se stejnými či různými rozptyly růstových parametrů. Také však může být pozorované rozložení růstových křivek způsobeno tím, že v populaci existuje několik skupin, které neznáme, v nichž se odehrává odlišný vývoj. Tyto skupiny mohou být různě velké a kromě toho, že se mohou lišit v průměrných vývojových trajektoriích se mohou lišit i míře individuálních rozdílů metu trajektoriemi uvnitř skupin. Některé skupiny tak mohou mít fixní vývojovou trajektorii, kdy jsou jakékoli odchylky od této trajektorie považovány za náhodnou chybu měření, zatímco jiné skupiny mohou být popsány tak, že jsou v nich vývojové

trajektorie, které mají nějaký průměr, od kterého se ostatní liší s nějakým rozptylem v normálním rozložení.

Z určitého úhlu pohledu je možné za příliš vysokou heterogenitu trajektorií považovat i situaci, kdy se rozptyl náhodných růstových parametrů signifikantně liší od nuly (Wickrama, Lee, O'Neal, & Lorenz, 2016, Jung & Wickrama, 2008). Z této perspektivy je pak cílem výzkumu identifikovat příčiny či prediktory rozdílnosti trajektorií, k čemuž se dají využít dostupné teoreticky odůvodněné kovariáty vysvětlující v modelu rozptyl růstových parametrů. Přesto může zůstat nevysvětlený nenulový rozptyl.

V situaci, kdy teorie nenabízí proměnné, které by vysvětlovaly rozdíly v trajektoriích, nebo pokud naopak teorie nabízí velké množství nikoli nezávislých vlivů, se nabízí myšlenka čistě empirické klasifikace. Lze tedy hledat subpopulace různých velikostí, které mají specifické typické (průměrné) vývojové trajektorie. Mezi subpopulacemi mohou být rozdíly ve formě/tvaru vývojových trajektorií, pouze v některých parametrech trajektorií, či ve variabilitě těchto parametrů. Jedním z přístupů, který umožňuje empirickou kategorizaci růstových křivek je *growth mixture modeling* (GMM), tedy modelování růstových směsí, vycházející z modelů konečných směsí normálních rozložení (Nagin, 1999, 2009) a příbuzné modely.

Podobně jako již v sociálních vědách dobře etablovaná *analýza latentních tříd* (latent class analysis, LCA, popř. latent profile analysis, LPA, např. Marsh, Lüdtke, Trautwein, & Morin, 2009) jde o přístup, který předpokládá, že pozorovaná heterogenita individuálních růstových křivek v populaci je výsledkem toho, že populace se skládá ze dvou či více subpopulací, uvnitř kterých můžeme růstové křivky považovat za homogenní (ať již ve smyslu normálně rozložené, nebo s nulovým rozptylem – konstantní). Tyto subpopulace mohou být reálně existující, nebo (a to patrně častěji) jen heuristickými nástroji zastupujícími komplexitu, kterou v rámci prováděné analýzy, probíhajícího výzkumu nemůžeme nebo nechceme podrobněji modelovat. Takovým způsobem postupovali například Luoma, Korhonen, Salmelin, Helminen, & Tamminen (2015) při identifikaci typů vývoje depresivních symptomů u matek. Kromě přístupu založeného na latentních třídách či modelech směsi lze uvažovat i aplikaci shlukové analýzy na růstové parametry. Tento přístup navrhovali například Dumenci & Windle (2001), ale příliš se neujal.

Modelování růstových směsí má poměrně dlouhou historii (Duncan, Duncan, & Strycker, 2006). Navazuje na pokroky v oblasti aplikace mixture přístupu na modely kovariančních struktur, specificky Yungův CFA mixture model (Yung, 1997) a mixture strukturních regresní model Armingera a Steinové (1997). Dalším výraznou inspirací byl Nagin (Nagin, 2009), který aplikoval LCA na růstové trajektorie a pojmenoval svůj model *semiparametric group-based trajectory model*. (Vermunt & Dijk, 2001) jej přejali pod názvem *latent-class regression model* a začlenili so svého software Latent Gold (Vermunt & Magidson, 2008). Bengt Muthén a jeho kolegové aplikovali tento přístup na latentní růstové modely a navrhli obecný rámec pro modelování růstových směsí (*general growth mixture modeling framework*, GGMM) a zároveň jej zpřístupnili badatelům tím, že jej začlenili do jimi vyvíjeného a hojně užívaného software Mplus (Muthén & Muthén, 2000, Muthén, 2001)

Cílem této práce je podobněji představit model růstových směsí a příbuzné modely, aplikovat je na dostupná data z longitudinálních výzkumů realizovaných na Institutu výzkumu dětí, mládeže a rodiny a zvážit jejich potenciál pro tento typ dat. Na řadu proměnných jsme v minulosti s kolegy aplikovali analýzu růstových křivek (např. Beranová, Ježek & Širůček, 2011, Ježek, Masopustová & Bouša, 2011), aniž bychom v datech našli

proměnné, které by uspokojivě vysvětlovaly variabilitu pozorovaných růstových křivek. Na tato data jsem nyní aplikoval model růstových směsí.

I. MODEL Y HETEROGENITY RŮSTOVÝCH KŘIVEK

Modely růstových křivek typicky předpokládají, že růstové parametry křivek (úroveň, tempo růstu) jsou v populaci normálně rozloženy. Mají tedy nějaký průměr, který může být použit pro určení střední, průměrné vývojové křivky v dané populaci, a rozptyl, který udává, jak moc se jednotlivci v populaci v daném růstovém parametru liší. Pokud je rozptyl růstových parametrů relativně malý, pak "průměrná" růstová křivka poměrně dobře reprezentuje populaci, lze o ní uvažovat jako o prototypické. Takový model se jeví snadno interpretovatelným. Příkladem by mohly být růstové křivky BMI dítěte (např. Cimino et al., 2016). V psychologii, s ohledem na charakter modelovaných proměnných a přesnost jejich měření, je však heterogenita růstových parametrů často tak vysoká, že zahrnuje jedince, u kterých sledovaná charakteristika roste i klesá a tento růst a pokles se odehrávají na různých úrovních rysu, kde pak mohou mít i odlišné konsekvence (je něco jiného, když 10leté dítě hubne z 80 kg, než když hubne z 30 kg). V takové situaci je pak porozumění heterogenitě (parametrů) růstových křivek hlavním cílem, jehož naplnění umožňuje interpretaci modelu a jeho možné další využití.

Modelování heterogenity může mít dva teoreticky velmi odlišné důvody, jak uvádějí (Bauer & Curran, 2003). Tím prvním, teoretickým, je identifikace subpopulací, jejichž vývoj sledované veličiny se kvalitativně liší. Druhým důvodem je modelování rozložení růstových parametrů, která nejsou normální, jejich kategorickou aproximací. Mezi těmito dvěma cíli je důležité rozlišovat, zejména při interpretaci tříd. Latentní třídou tak můžeme mýnit jak skutečně existující subpopulaci, kterou se zatím nepodařilo identifikovat a popsat, tak pouze ad-hoc identifikovanou skupinu, která usnadňuje analýzu a tvorbu teorie.

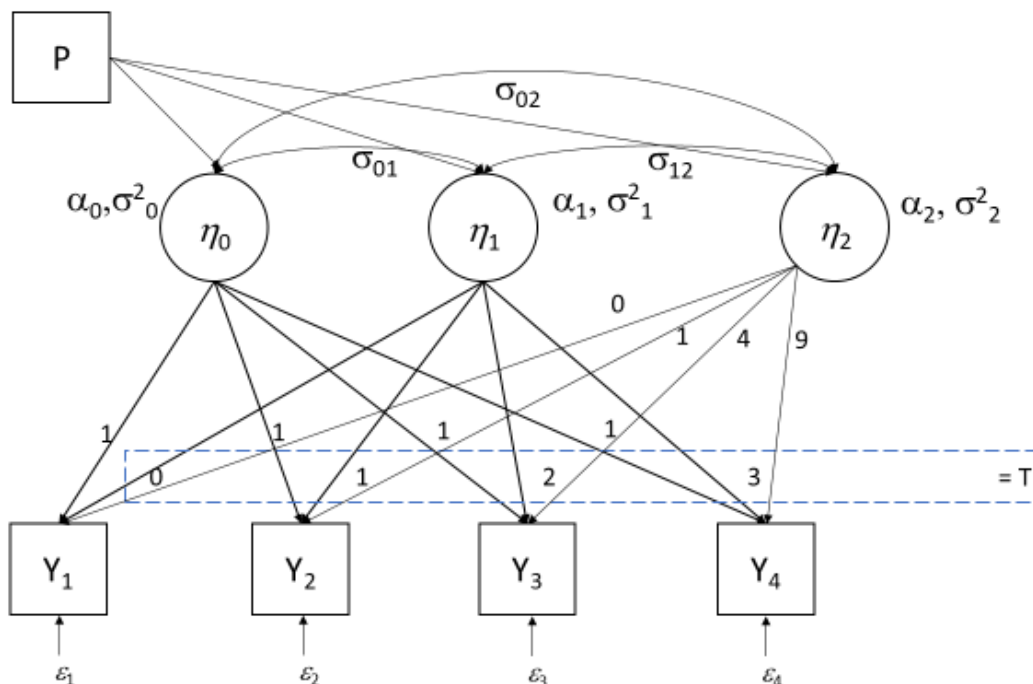
Analýza latentních růstových křivek a model růstových křivek

Výchozí variantou je heterogenitu akceptovat v té podobě, v jaké je modelována v modelu růstových křivek (GCM) či latentních růstových křivek (LGMA). Tedy s ohledem na teorii přijmout to, že vývoj charakteristiky je velmi rozmanitý, přičemž čím vzdálenější je růstová křivka od střední růstové křivky, tím je méně pravděpodobná, popř. v populaci méně prevalentní. Rozložení růstových křivek je spojitě a teorie nedává důvod k nějaké kategorizaci. Nepodmíněný normální kvadratický model růstu lze v rámci latentních růstových modelů dle Kreuter & Muthén (2008) zapsat následovně:

$$\begin{aligned} Y_{ij} &= \eta_{0i} + \eta_{1i}T_j + \eta_{2i}T_j^2 + \varepsilon_{ij} , \\ \eta_{0i} &= \alpha_0 + \xi_{0i} , \\ \eta_{1i} &= \alpha_1 + \xi_{1i} , \\ \eta_{2i} &= \alpha_2 + \xi_{2i} , \end{aligned}$$

kde η jsou náhodné růstové parametry s populačními průměry α . Individuální odchylky od těchto průměrných hodnot populačních parametrů ξ pak mají v rámci tohoto modelu normální rozložení s rozptyly σ^2 - právě tyto rozptyly zde reprezentují heterogenitu růstových křivek. Tuto heterogenitu se můžeme pokusit vysvětlit, predikovat pomocí spojitých i kategorických prediktorů, které mohou být do modelu zařazeny a mohou predikovat kterýkoli z náhodných růstových parametrů. V podobě strukturního modelu, který je až na prediktor P ekvivalentní nepodmíněnému modelu zapsanému víceúrovňovými regresními rovnicemi výše, je to znázorněno na Obrázku I.1.

Obrázek I.1. Kvadratický model latentních růstových křivek s manifestním prediktorem parametrů růstu.



Analýza latentních tříd růstových křivek

Jinak budeme k heterogenitě přistupovat, pokud existují důvody domnívat se, že rozložení růstových křivek není spojitě, že odlišnosti pozorovaných individuálních růstových trajektorií jsou výsledkem toho, že v populaci je několik skupin, které se liší svým vývojem, přičemž uvnitř skupin se všichni jejich členové vyvíjejí shodně. Zdánlivá spojitost rozložení růstových křivek je pak připisována jen chybě měření. Z tohoto předpokladu vychází modely *latentních tříd růstových křivek* (latent class growth analysis, LCGA, Muthén & Muthén, 2000) známé též jako modely skupinových trajektorií (group-based trajectory models, Nagin, 1999). Model latentních růstových tříd vycházející z výše uvedeného nepodmíněného normálního kvadratického modelu růstu bychom lze zapsat následovně:

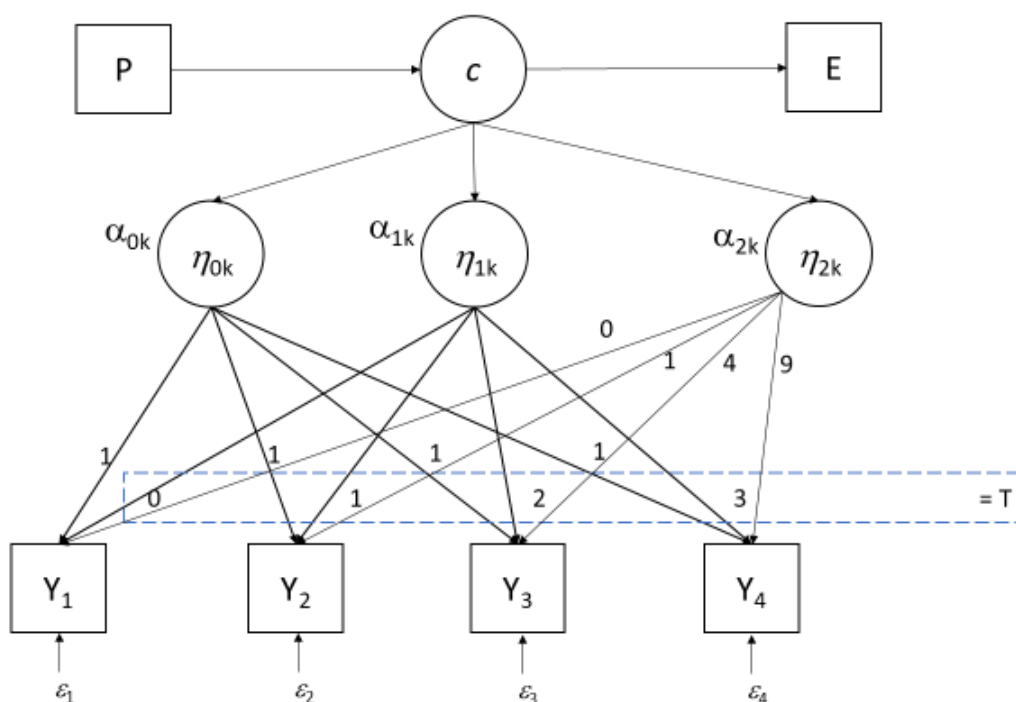
$$\begin{aligned}
 Y_{ij|c_i=k} &= \eta_{0k} + \eta_{1k}T_j + \eta_{2k}T_j^2 + \varepsilon_{ij}, \\
 \eta_{0k} &= \alpha_{0k}, \\
 \eta_{1k} &= \alpha_{1k}, \\
 \eta_{2k} &= \alpha_{2k},
 \end{aligned}$$

kde k reprezentuje číslo latentní třídy, do které jednatel s pravděpodobností implikovanou modelem spadá. Každá skupina k tak má svou hodnotu růstových parametrů α_0 - α_2 , a ta je v rámci skupiny konstantní. Jakékoli odchylky se tak stávají součástí ε_{ij} .

Je-li cílem dále analyzovat (predikovat, hledat kovariáty) parametry růstových křivek, máme při použití LCGA jedinou možnost, a to zařadit prediktory či efekty latentní třídy, protože růstové parametry samotné jsou uvnitř latentních tříd konstantní. Model latentních tříd kvadratických normálních růstových křivek s jedním prediktorem členství v třídě a

jedním efektem členství v třídě je v podobě strukturního modelu zobrazen v Obrázku I.2. Latentní kategorická proměnná reprezentující členství ve třídě je zde označena c .

Obrázek I.2. Model latentních tříd růstových křivek s prediktorem třídy a efektem třídy.



Stejně jako u latent-class či latent-profile analysis obvykle nevíme z teorie, kolik subpopulací v populaci existuje a jak jsou velké. Výsledkem analýzy pak je argumentace pro určitý počet skupin (tříd) s jejich specifickými růstovými trajektoriemi a pravděpodobnostní členství jednotlivců v těchto skupinách, něhož můžeme usuzovat i na velikost skupin. Otázkou zde je, nakolik považovat výsledné latentní třídy za skutečné a nakolik za analytický artefakt, heuristickou pomůcku, či aproximaci. Raudenbush (2005, cit dle Kreuter & Muthén, 2008) dochází k závěru, že bychom měli vycházet z předpokladu spojitě variability růstových trajektorií, a pouze jasné teoretické důvody by nás měly vést ke kategorizujícímu přístupu.

Nezapomínejme však, že typickým, a ne zcela snadno změnitelným předpokladem GCM/LGMA je normální rozložení parametrů. Víme-li z dat, že rozložení parametrů je výrazně nenormální, jeví se účelné pokoušet se popsat toto rozložení aproximující kategorizací – tedy latentními třídami růstových křivek – která je navíc relativně parametricky úsporná.

Růstový mixture model

Přístupem, který kombinuje výše uvedené, jsou růstové mixture modely (growth mixture models, GMM, Muthén & Muthén, 2000; Ram & Grimm, 2009). V růstovém mixture modelu modelujeme také latentní třídy¹ s jejich typickými růstovými trajektoriemi. Růstové parametry však nejsou uvnitř latentních tříd konstantní – jsou náhodné a mají normální rozložení s nenulovým rozptylem. Celková heterogenita růstových křivek je tak modelována jako směs

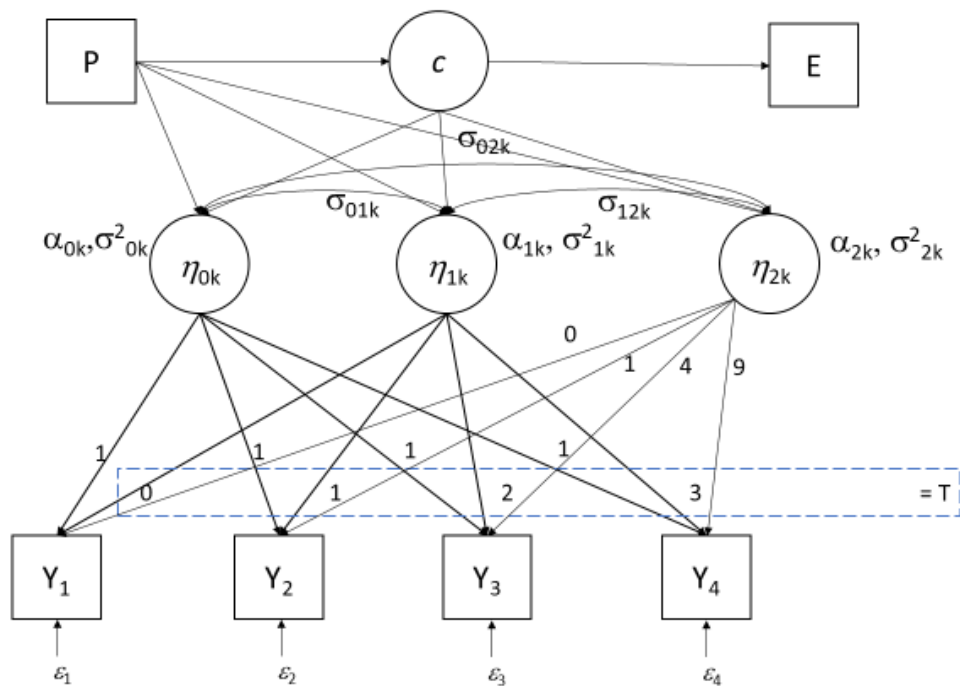
¹ V tradici mixture modelů (modelů směsí) se místo termínu latentní třída používá spíše komponenta směsi.

spojitých pravděpodobnostních rozložení. Normální kvadratický růstový mixture model vycházející z výše uvedeného nepodmíněného normálního kvadratického modelu růstu lze zapsat následovně:

$$\begin{aligned}
 Y_{ij|c_i=k} &= \eta_{0ki} + \eta_{1ki}T_j + \eta_{2ki}T_j^2 + \varepsilon_{ij} , \\
 \eta_{0ki} &= \alpha_{0k} + \xi_{0ki} , \\
 \eta_{1ki} &= \alpha_{1k} + \xi_{1ki} , \\
 \eta_{2ki} &= \alpha_{2k} + \xi_{2ki} .
 \end{aligned}$$

Toto řešení je flexibilnější než výše uvedené možnosti, avšak kvůli velkému množství parametrů méně stabilní, náročnější na množství dat a hůře identifikovatelné. V obrázku I.3, který zobrazuje růstový mixture model s kovariátou, si lze povšimnout, že subscripty k přibýly ke všem parametrům růstových křivek - k průměrným hodnotám, rozptylům i kovariancím. To znamená, že s rostoucím počtem tříd roste v plně specifikovaném GMM modelu počet odhadovaných parametrů velmi rychle. Proto např. Mplus v tomto modelu defaultně fixuje rozptyly a kovariance růstových parametrů na stejnou hodnotu napříč skupinami.

Obrázek I.3. Kvadratický růstový mixture model s prediktorem a efektem latentní třídy.



V růstovém mixture modelu mohou prediktory/kovariáty vývoje predikovat jak členství v latentní třídě, tak variabilitu růstových parametrů uvnitř třídy.

Jak uvádí Kreuter a Muthén (2008), předpoklad normality rozložení náhodných parametrů je uvnitř tříd spíše naplněn než u celé populace. Přesto je někdy toto rozložení výrazně nenormální a je třeba hledat další cesty, jak variabilitu vývojových trajektorií modelovat. Jednou z možností je neparametrický GMM (Muthén, Asparouhov, 2008). Rozložení náhodných parametrů uvnitř tříd je v něm aproximováno pomocí "subtříd". "Průměrné" trajektorie uvnitř tříd se nezmění, pouze se k nim přidá informace o Dk subtřídách, které se budou lišit v náhodných parametrech (uvnitř subtříd jsou parametry konstantami).

Růstové parametry a jejich zobrazování

Pro účely interpretace a vizualizace je vhodné si umět představit, jak se parametry modelu projevují v rozpětí křivek popsané třídou, resp. komponentou směsi. V publikovaných mixture analýzách se lze běžně setkat s velmi elaborovanými modely, které jsou však graficky prezentované tím nejjednodušším možným způsobem, tedy grafem průměrné růstové křivky v každé jednotlivé třídě (viz např. Robinson, Perez, Nuttall, Roseth, & Linnenbrink-Garcia, 2018). To snadno vede k nadhodnocené představě o rozdílech mezi latentními třídami, či o jejich separaci.

Již u běžného modelu latentních růstových křivek s jednou třídou není vždy snadné si představit, jaké případy individuálního růstu model dobře reprezentuje, a které bychom z hlediska modelu vnímali spíše jako outliery nebo i chyby v datech. V případě triviálního modelu latentních průsečíků je to relativně snadné, protože růstové křivky jsou paralelní a liší se pouze svým umístěním na dimenzi modelované charakteristiky (ose Y). Rozpětí paralelních křivek je dáno rozptylem latentních průsečíků, což je přímo parametr modelu. Není tedy obtížné si představit střední růstovou křivku a nad ní a pod ní paralelní křivky vzdálené 1 či 2 směrodatné odchylky průsečíku. Tím lze získat představu o rozpětí růstových křivek. Není to však ještě úplný obrázek dat, které model popisuje, protože takový graf neobsahuje reziduální rozptyl, tedy náhodné odchylky naměřených hodnot od jednotlivcovy individuální růstové křivky.

Pokud model růstových křivek obsahuje více náhodných parametrů křivek než průsečík a mezi těmito parametry jsou nenulové kovariance, je poměrně obtížné si představovat rozpětí poloh a tvarů růstových křivek popsaných modelem. V takovém případě se zaměřujeme na modelem implikovaný průměr a rozptyl sledované charakteristiky v různých časech (obvykle v časech měření zahrnutých v designu studie), přičemž individuální růstové křivky si představujeme v rámci pásma vytvořeného zhruba jednu či dvě SD nad a pod průměrnou hodnotou v každém čase měření. To však poněkud zakrývá možné individuální tvary individuálních růstových křivek, a proto můžeme interpretační vizualizaci toho, co model reprezentuje, vytvořit také mnohonásobným náhodným losováním parametrů podle kovarianční matice růstových parametrů a vykreslením jim odpovídajících růstových křivek (ať již s přičtením náhodného rozptylu nebo bez). Příklady takových grafů jsou v kapitolách II. a III. pro růsty užívání návykových látek a prožívání autonomie.

Když od modelu růstových křivek pokročíme k modelu latentních tříd růstových křivek nebo růstovému mixture modelu, situace se ještě zkomplikuje. Na jednu stranu bychom mohli každou jednotlivou latentní třídu vykreslit samostatně stejně jako u LGCM. Tak bychom získali představu o tom, jak by vypadaly růstové křivky v jednotlivých latentních třídách. To je pro porozumění a interpretaci toho, v čem se růst v různých latentních třídách liší. Kdybychom však chtěli takovou vizualizaci porovnat se zobrazením hrubých dat, nemusely by se tato zobrazení nutně dobře překrývat tak, abychom měli dobrý pocit, že model i vizuálně odpovídá datům. Důvodem je to, že členství v třídách je v mixture modelech pravděpodobnostní, a i v případě poměrně vysoké oddělenosti tříd a tím i vysokých pravděpodobností členství v nejpravděpodobnější třídě (entropie) jsou v datech obtížně kategorizovatelné případy, které mají podobnou pravděpodobnost členství ve všech modelovaných třídách. Pokud tedy chceme vykreslit, jaké růstové křivky lze na základě modelu v populaci očekávat, je potřeba v odhadu průměru a rozptylu jednotlivých indikátorů latentní třídy zohlednit i klasifikační nejistotu. Protože to již je matematicky poměrně komplikované, vychází Mplus uživatelům vstříc a pomocí příkazu TECH7; vypisuje odhad

výběrových statistik manifestních proměnných vážených pravděpodobnostním členstvím v latentních třídách (Muthén & Muthén, 2017). Vykreslení této variability již Mplus neposkytuje, a proto je potřeba grafy s rozpětími křivek vytvořit v pomoci jiného software, např. R.

Růstové modely vyšších řádů

Modely latentních tříd růstových křivek a růstové mixture modely je možné poměrně přímočaře rozšířit na modely růstu latentních proměnných, tzv. *růstové modely druhého řádu* (second-order/higher-order latent growth models, Wickrama et al., 2016). I když jsou totiž tyto modely z psychometrického hlediska velmi žádoucí a poměrně dlouho známé (psychologům je představil již McArdle, 1988), v publikované empirické literatuře se ani dnes příliš nevyskytují (Geiser, Keller, & Lockhart, 2013).

Růstové modely vyšších řádů se vyskytují pod mnoha názvy a zkratkami (SO-LGM, Wickrama et al., 2016, SGM, Geiser et al., 2013) a jejich konkrétní podoby pod ještě dalšími názvy, takže nemusí být na první pohled patrné, že jde o jednu rodinu modelů. Základním, a z hlediska psychometrie nejatraktivnějším, rysem těchto modelů je to, že analyzují vývoj nikoli manifestní proměnné, součtového skóru reprezentujícího zkoumaný konstrukt, ale latentního skóru, faktoru spolu s jeho modelem měření. Z toho plynou následující výhody, jak je shrnují Geiser et al. (2013, s. 480):

- oddělují chybu měření od růstu či změny rysu a reliabilního časově specifického rozptylu, čímž umožňují rozlišení vývoje rysu od stavu ve smyslu latent state-trait teorie,
- umožňují testovat předpoklad longitudinální invariance měření,
- mají větší statistickou sílu pro detekci individuálních rozdílů v růstu,
- díky práci s více indikátory konstruktů (položkami, subtesty, metodami) umožňují oddělit konstruktový rozptyl od specifického metodového rozptylu.

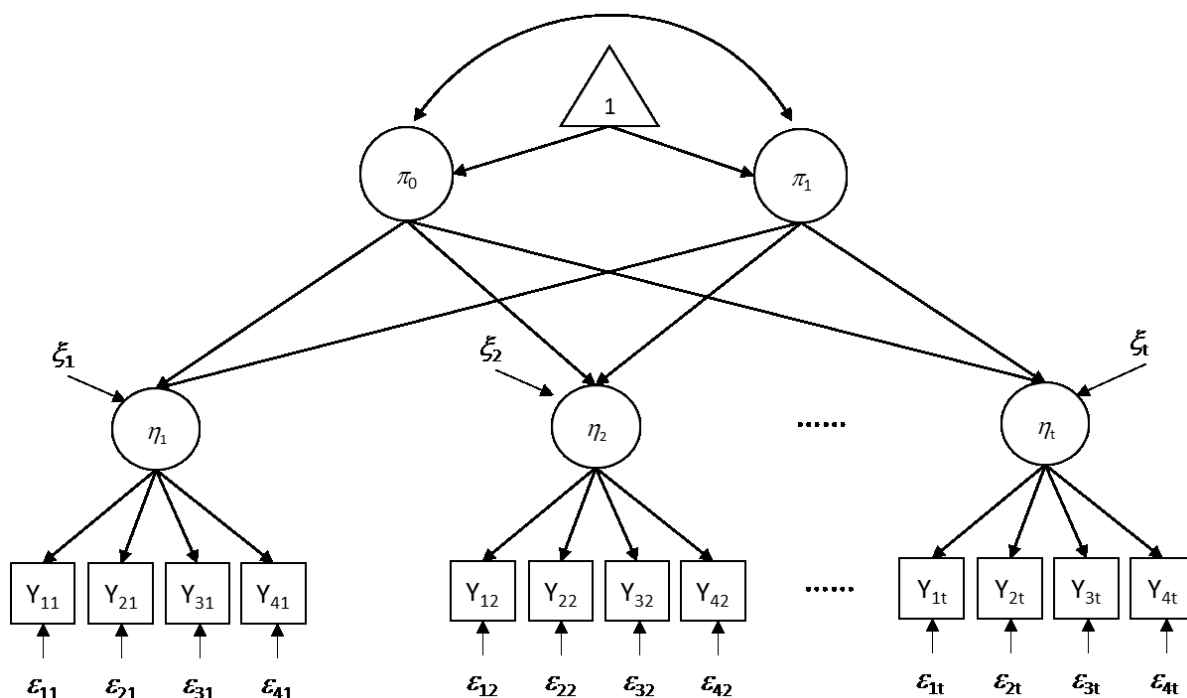
Růstové modely vyšších řádů tak dávají možnost se mnohem lépe ujistit, že modelujeme právě a pouze vývoj konstruktů, o který se zajímáme, a ne různých náhodných či stabilních rušivých vlivů, které se podílejí na omezené validitě a reliabilitě měření. Dávají možnost ověřovat předpoklady, které platí i v růstových modelech prvního řádu, kde však často vůbec nejsou ani vyslovovány, protože tam není způsob, jak posoudit jejich platnost. Konkrétně invariance měření je takovým obvykle nevyřčeným předpokladem. I v běžných GCM či LGMA modelech se předpokládá, že proměnná, jejíž vývoj modelujeme, reprezentuje ve všech časech měření konstrukt našeho zájmu stejně, že použité měřítko měří stále totéž, stejně vyvážený mix facet konstruktů. Tento předpoklad však nemusí být nutně naplněn – s věkem či s opakovaným použitím měřítka se může proměňovat, jak funguje, jak respondenti rozumí položkám, či hodnotí svou pozici na položkách (Meredith & Horn, 2001). Prostřednictvím simulovaných změn v modelu měření zjistili (Cole, Bauer, Hussong, & Giordano, 2017), že i malé změny v modelu měření ovlivňují jak enumeraci latentních tříd, tak jejich parametry v analýze latentních tříd a factor mixture modelu (Lubke & Muthén, 2005).

RŮSTOVÝ MODEL FAKTORŮ

Nejpřímochařejším zobecněním LGCA modelu je model nazývaný *curve of factor scores* model (CUFFS, McArdle, 1988) nebo *curve of factors model* (CFM, Wickrama et al., 2016, Isiordia & Ferrer, 2018). Ve zjednodušené podobě je znázorněn na obrázku I.4. Namísto toho, aby latentní průsečík π_0 a latentní směrnice π_1 determinovaly vývoj přímo manifestního měřítka zkoumaného konstrukt, determinují vývoj skóre latentní proměnné η_1 - η_t . Ta reprezentuje zkoumaný konstrukt prostřednictvím modelu měření, v němž je latentní skóre specifikovaný jako společný faktor položek Y_1 - Y_4 . Model měření je ve všech časech identický, což zaručuje srovnatelnost latentních skóre napříč časy. V tomto modelu parametry latentní růstové přímky π_0 a π_1 determinují určitou část rozptylu latentních skóre - to je systematický rysový rozptyl. Zbývající část rozptylu latentních skóre (disturbance ξ_1 - ξ_t) je reliabilní situační rozptyl měřeného konstrukt - mohli bychom říci, že to je stavová komponenta zkoumaného konstrukt. Konečně rezidua jednotlivých položek (ε_{11} - ε_{4t}) reprezentují z hlediska měřeného konstrukt náhodný či irelevantní rozptyl.

Růstovou část modelu lze samozřejmě specifikovat i složitěji, pokud je třeba. Může být i kvadratický s dalším latentním růstovým členem.

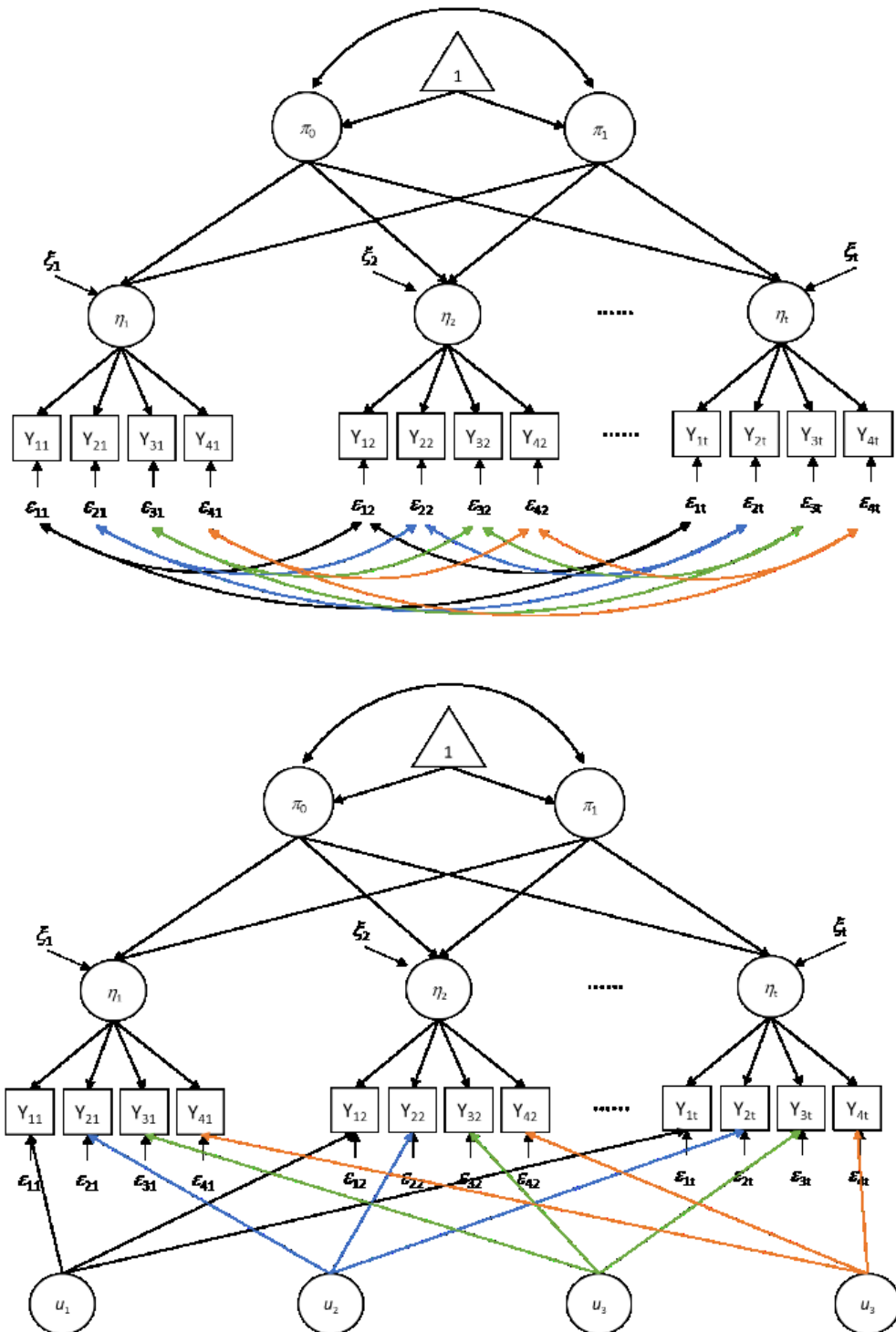
Obrázek I.4. Růstový model faktorů (CFM).



Vzhledem k tomu, že v modelu jsou dvojí reziduální rozptyly - disturbance a reziduální rozptyly položek - je potřeba stanovit dvě reziduální struktury - strukturu disturbance a strukturu reziduálních rozptylů položek. Pro rezidua položek je běžným předpokladem, že vedle náhodného rozptylu obsahují jedinečný položkový rozptyl, a proto je vhodné nechat rezidua též položky napříč časy korelovat. Pro tyto longitudinální korelace reziduí je možné zvolit i parametricky úspornější struktury, jako jsou AR, Toeplitz. Při větším počtu měření je možné namísto korelační struktury specifikovat položkový/metodový faktor, který sytí všechny instance též položky/metody - jeden pro každou položku (Wickrama et al., 2016). (Geiser & Lockhart, 2012) uvádí, že i když může specifikace položkových faktorů pomoci

osvětlit faktorovou strukturu měřícího nástroje, je na místě je použít až po ověření longitudinální invariance a ověření dostatečné velikosti longitudinálních korelací mezi rezidui - ty by měly mít alespoň střední velikost.

Obrázek I.5. CFM s autokorelační strukturou reziduí a CFM s položkovými faktory



Podobně struktura disturbancí může mít širokou paletu podob od nekorelovaných disturbancí, kdy stavová komponenta sledované charakteristiky nemá žádnou setrvačnost, přes jednoduchou strukturu korelací mezi sousedními měřeními, až po složitější korelační struktury.

Z hlediska interpretace růstových parametrů růstového modelu faktorů se zde oproti běžnému LGC modelu mnoho nemění. Latentní průsečík a směrnice s jejich průměry, rozptyly a kovariancí nesou stejný význam a popisují rozložení individuálních růstových křivek – tentokrát však reliabilního faktorového skóru sledované veličiny.

Stejně jako u LCGA modelů lze heterogenitu růstových křivek modelovat pomocí kovariátů, popř. pomocí kovariátů vysvětlovat časově specifický rozptyl sledované veličiny, nebo aplikovat analýzu latentních tříd, či mixture model pro hledání kategorické proměnné vysvětlující rozdíly mezi individuálními růstovými křivkami. Příkladem je nedávná studie (Robinson et al., 2018) modelující vývoj přijímání identity přírodního vědce u univerzitních studentů.

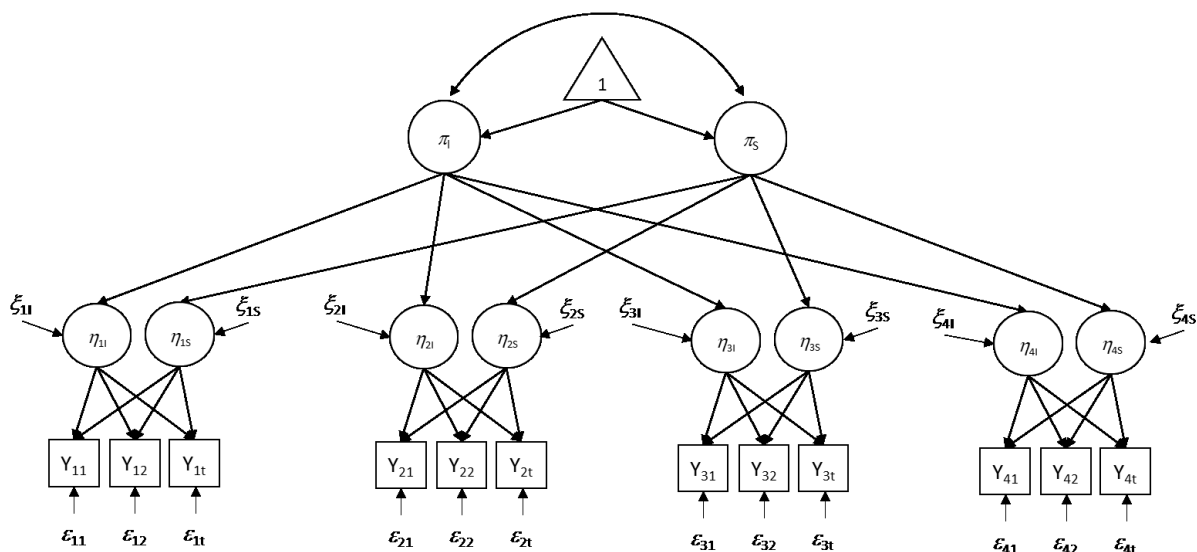
MODEL FAKTORŮ KŘIVEK

Konceptuální alternativou k růstovému modelu faktorů je model faktorů křivek – *factor of curves model* (FCM, Wickrama et al., 2016) či FOCUS (McArdle, 1988). Namísto specifikování růstového modelu nad latentními měřítky je růstový model specifikován pro každou opakovanou položku/indikátor a latentní proměnné vyššího řádu pak specifikují vztahy mezi růstovými parametry jednotlivých položek. Pokud všechny položky či indikátory měří tentýž vyvíjející konstrukt měly by být jejich růstové parametry vysoce korelované. Všechny latentní průsečíky by tak mohly být vysoce syceny latentním průsečíkem vyššího řádu podobně jako všechny latentní směrnice latentní směrnici vyššího řádu. Zjednodušený diagram modelu faktorů křivek pro čtyřpoložkovou škálu a tři měření je na obrázku I.6.

Model faktorů křivek je vlastně zobecněním modelu paralelních procesů (PPM, *parallel-process model* či *associative LGM*, Muthén & Curran, 1997), který uvádí do vztahu parametry dvou růstových křivek modelujících dva jevy. Ve FCM nejde o dva příbuzné jevy, ale více položek či indikátorů jevu, který je předmětem našeho zájmu. Na rozdíl od růstového modelu faktorů (CFM) zde není součástí modelu model měření. To je z hlediska v psychologii převládajícího reflektivního měření zjevná nevýhoda. Na druhou stranu, pokud jde o formativní měření, v němž dávají položky vzniknout součtovému skóru, byl by CFM misspecifikací. Model faktor křivek je tak vhodný tehdy, když modelujeme vývoj indexů, nebo když modelujeme vývoj jevu, který je reprezentován několika samostatnými konstrukty. Například (Wickrama et al., 2016) popisují model, kde je míra patologie modelována pomocí měřítek depresivity a úzkosti.

I u tohoto modelu je potřeba uvážlivě specifikovat reziduální strukturu. Nenulové korelace disturbancí ukazují, že dvě položky se vyvíjejí společně podobněji, než by implikoval společný průsečík či směrnice druhého řádu. Z této perspektivy jsou tyto korelace nežádoucím, rušivým prvkem modelu a je dobré otestovat, zda jejich fixování na 0 zhoršuje shodu modelu s daty. Kovarianční struktura položkových reziduálních rozptylů může představovat výzvu. Je totiž potřeba zvážit jak kovariance v rámci růstových modelů jednotlivých položek, které mohou mít všechny podoby uvedené v kapitole o latentních růstových modelech, tak kovariance reziduálních rozptylů různých položek v témže čase – ty by reprezentovaly časově-specifický rozptyl modelovaného jevu. Takto složitá struktura může snadno narazit na problémy s identifikací a konvergencí.

Obrázek I.6. Model faktorů křivek



Další příbuzné modely

Konfirmační modely s latentními třídami. Finch a Bronk (2011) referují o konfirmační LCA. Ta je analogií CFA. Jde tedy o testování hypotézy o tom, že v populaci je určitý počet tříd a tyto třídy mají určité charakteristiky. Do určité omezené míry jsou konfirmační prvky součástí obvyklého průběhu modelování latentních tříd či směsí, protože v souladu s teoretickými očekáváními se uvolňují či omezují parametry tříd a také se volí počet tříd. V podání Finchea a Bronk (2011) jde tedy spíše o zdůraznění nutnosti konfirmačního uvažování. V první řadě navrhují nastavení startovacích hodnot parametrů tříd (ne velikosti tříd) na hodnoty očekávané teorií. Algoritmus odhadu se ovšem od startovacích hodnot může libovolně vzdálit, a tak je posouzení míry shody odhadnutých parametrů s očekávanými (startovacími) ryze subjektivní. Finch a Bronk (2011) zvažují i pevné fixování alespoň některých parametrů tříd. To však má poměrně zásadní dopad na ukazatele shody modelu s daty (jako bychom pevně fixovali hypotetizované náboje položek v CFA), a pokud nefixujeme všechny parametry tříd může to velmi ovlivnit podobu tříd s nezafixovanými parametry. Patrně konfirmačně nejužitečnější technikou je zde využití omezení parametrů v podobě nerovností (pomocí MODEL CONSTRAINTS). Ty mohou být využity i u modelů směsí, a tedy i u GMM modelů. Průměr (pravděpodobnost) indikátoru třídy může být omezen tak, aby byl v jedné třídě vyšší než v jiné třídě, popř. aby byl násobkem (např. -1). Tak lze dosáhnout dostatečného omezení volnosti parametrů reflektující teorii, aby pak bylo smysluplné porovnávat shodu takto omezeného modelu s daty se shodou modelu neomezeného. Konfirmační postup použili například Donovan a Chung (2015) pro klasifikaci užívání alkoholu v adolescenci.

Modely s latentními třídami či modely směsí na víceúrovňových datech. Stejně jako u běžných lineárních modelů předpokládají výše popsané modely lokální nezávislost reziduí, tedy to, že při zohlednění všech vztahů mezi proměnnými explicitně zahrnutých do modelu či implikovaných modelem, jsou reziduální rozptyly nezávislé. Nejčastější situací, v níž je tato podmínka ohrožena či narušena, jsou víceúrovňová data, tedy data získaná na vzorcích zahrnující vnořené klastry účastníků. I pro tato data existují rozšíření modelu latentních tříd – MLCA modely. Svá řešení a implementace nabízí jak (Vermunt, 2008) za Latent GOLD, tak (Muthén & Asparouhov, 2009) za Mplus. Použití víceúrovňových modelů s latentními třídami

vede k přesnějším odhadům parametrů, přesnějším směrodatným chybám a spolu s tím i přesnějším testům hypotéz o jednotlivých parametrech i celém modelu. Různé způsoby parametrizace MLCA modelů porovnávají (Finch & French, 2014). Takové modely pak nabízejí řešení zajímavých otázek, například Van Horn et al. (2016) modelují, jak latentní třídy na úrovni jednotlivců (úroveň 1) mohou moderovat vztahy mezi proměnnými na vyšší úrovni.

Dynamická LCA. Asparouhov, Hamaker, & Muthén (2017) nedávno představili zcela odlišný rámec pro analýzu intenzivních longitudinálních dat, tedy dat s velkým množstvím opakovaných měření. Model zachycuje vývoj každého jedince jako postupně měnící se členství v latentních třídách, přičemž změny mezi jednotlivými měřeními jsou považovány za Markovský proces.

Přístupem, který kombinuje analýzu latentních tříd s principy tradiční hierarchické klastrové analýzy, jsou *latent class growth trees* (van den Bergh, Schmittmann, & Vermunt, 2017). Tento přístup nepředpokládá, že subpopulace, třídy jsou latentní a podle prvních okusů o jejich užití se zdají být perspektivní alternativou mixture modelů (Jacobucci, Grimm, & McArdle, 2017). Konečně, modely s latentními třídami mohou pracovat s více než jednou latentní kategorickou proměnnou, jak ukazují Akushevich, Kovtun, Manton, & Yashin (2009).

II. ZÁKLADNÍ POSTUP ANALÝZ HLEDAJÍCÍ LATENTNÍ TŘÍDY

Základními cíli analýzy s latentními třídami je (a) určit počet latentních tříd ve směsi, (b) odhadnout rozložení indikátorů pro každou latentní třídu - podmíněné pravděpodobnosti pro kategorické indikátory, podmíněné rozložení pro spojité indikátory a (c) stanovit pravděpodobnostní členství jednotlivců v latentních třídách (Duncan et al., 2006).

Podobně jako u jiných exploračně zaměřených analýz je identifikování latentních tříd vývojových křivek postup skládající se z řady kroků, z nichž se postupně vynořují možná řešení, která splňují jak teoreticko-interpretací nároky, tak nároky kladené na shodu modelů s daty a robustnost odhadů jejich parametrů. Základní postup se skládá z následujících kroků (Wickrama et al., 2016):

1. Specifikace výchozího nepodmíněného modelu latentních růstových křivek (LGCM)
2. Provedení analýzy latentních tříd růstových křivek (LCGA)
3. Specifikace růstového mixture modelu (GMM), popř. různě omezených GMM modelů
4. Řešení problémů s odhadem parametrů modelů
5. Volba optimálního nepodmíněného modelu s třídami
6. Zařazení kovariátů

Postup není jednosměrný a je běžné, že se v rámci analýzy vracíme k přehodnocení voleb učiněných v předchozích krocích.

1. Specifikace nepodmíněného modelu latentních růstových křivek

Prvním krokem je specifikace nepodmíněného růstového modelu. Oproti růstovému modelování, které nepředjímá možnost následného modelování heterogenity vývojových trajektorií pomocí latentních tříd, je zde potřeba uvažovat při specifikaci výchozího modelu poněkud jinak a zvažovat více možností. Snaha teoreticky zdůvodněně specifikovat model, který bude mít co nejlepší shodu s daty, může vést ke složitému modelu s mnoha parametry, v němž už ale s ohledem na jeho složitost bude velmi obtížné hledat latentní třídy, až již pomocí LCGA či GMM. Pomocí směsí více růstových tříd je možné modelovat i velmi složitý tvar a variabilitu růstových křivek, byť by růstový model jednotlivých latentních tříd byl velmi jednoduchý. Jde v zásadě o teoretické rozhodnutí badatele, zda se rozhodne pozorovaná data popisovat jednou složitou růstovou křivkou či více jednoduššími. Pro modelování heterogenity růstových trajektorií pomocí latentních růstových tříd tedy preferujeme spíše jednodušší výchozí model latentních růstových křivek. Důvodem pro preferenci jednoduchosti je prudký nárůst počtu parametrů s přibývajícimi latentními třídami komplikující získání jejich stabilních a nezkreslených odhadů.

U různých výchozích latentních růstových modelů je směrodatná především variabilita růstových parametrů - zda je zde vůbec přítomna heterogenita, kterou by bylo potřeba modelovat. To vůbec není triviální otázka, protože v psychologii není nijak výjimečné, že i přes velkou variabilitu modelované charakteristiky jsou si růstové křivky jednotlivců velmi podobné. To pak znamená, že většina variability zůstává v reziduálním rozptylu - pak vlastně model nepopisuje vývoj, ale proměňování proměnné v důsledku proměnlivých vlivů. Pokud je charakteristika silně ovlivňovaná vnějšími vlivy, a přesto potřebujeme zjistit, jestli se za tou variabilitou skrývá nějaký vývoj, pak je zde možnost identifikovat a změřit nejsilnější z těchto

vlivů a zařadit je do modelu. Pak lze a modelovat charakteristiku statisticky očištěnou od tohoto vlivu.

Rodina modelů růstových křivek, z nichž si svůj nepodmíněný model můžeme volit, je velmi rozsáhlá a každý prototypický model má řadu alternativních parametrizací, které mají dopad na to, jak budou v důsledku modelovány růstové parametry, jejichž náhodné rozdělení bude v dalších krocích analyzováno jako směs rozdělení. Můžeme volit mezi plně specifikovanými polynomickými modely, zejména lineárním a kvadratickým, ale lze použít i jejich varianty, kdy je tvar růstu specifikován jen zčásti (*level-and-shape*, *latent slope*, *latent base*), popř. použít více růstových parametrů (*piecewise*). Pro sigmoidní růsty jsou vhodné parametricky úsporné nelineární modely jako je logistický, Gompertzův či Richardsův (Grimm & Ram, 2009)

SPECIFIKACE RŮSTOVÉ (STRUKTURNÍ) ČÁSTI MODELU

Nepodmíněný LGC model modeluje růst tak, že stanovuje regresní rovnici pro hodnotu modelované charakteristiky každého jednotlivce (i) v každém čase měření (t).

$$Y(t)_i = \eta_{0i} + \eta_{1i}T(t) + u(t)_i,$$

kde $Y(t)_i$ je hodnota jednotlivce i v čase t ,
 $T(t)$ je hodnota proměnné udávající čas v čase t , někdy nazývaná **časová báze** (time base),
 $u(t)_i$ je reziduum jednotlivce i v čase t a
 η_{0i} a η_{1i} jsou individuální průsečík a směrnice

Nepodmíněnost modelu znamená, že parametry růstové křivky η_{0i} a η_{1i} , které mají napříč jednotlivci rozložení s nějakým průměrem a rozptylem, nejsou v modelu predikovány žádnými kovariáty. Součástí LGC modelu je předpoklad, že rozložení obou těchto parametrů je multivariační normální.

Individuální parametry η_{0i} a η_{1i} jsou v LGC modelu reprezentovány endogenními latentními proměnnými (proto jsou značeny η), které determinují (\rightarrow) manifestní hodnoty Y v jednotlivých časech měření. *Latentní průsečík* η_0 má v modelu fixovanou váhu 1 pro všechny manifestní měření Y . Pro různé jednotlivce nabývá tato latentní proměnná různé hodnoty, které bychom mohli dosadit do rovnice výše. Latentní směrnici η_1 také fixujeme její váhy – časové báze -- a to na hodnoty $T(t)$, tedy na hodnoty reprezentující vzdálenost jednotlivých měření v čase. Tyto hodnoty se nejčastěji nastavují tak, že hodnota prvního měření $T(1)$ je nastavena na 0 a hodnoty následujících časů měření jsou nastaveny tak, aby vyjadřovaly vzdálenost v čase od prvního měření. Pokud se měření odehrávala ve stejných intervalech, je běžné $T(t)$ kódovat 0, 1, 2, 3... j kde ($j+1$) je celkový počet měření. V takovém případě pak hodnota latentní směrnice η_{1i} pro jednotlivce vyjadřuje nárůst Y za jednotku času, kdy jednotkou času je čas uplynulý mezi dvěma měřeními. Toto nastavení metriky času je do značné míry arbitrární a jeho flexibilita je mocným nástrojem specifikace tvaru individuálních křivek. Než představím tuto paletu možností, je potřeba uvést, že pro zajištění identifikace všech parametrů modelu je nutné, aby u latentního průsečíku i latentní směrnice byla nakonec vždy alespoň jedna váha fixována a alespoň jedna váha byla fixována na hodnotu 0 (McArdle & Nesselroade, 2014, s. 94). I když 0 může být u kterékoli z obou latentních proměnných, obvykle to bývá u latentní směrnice. Latentní průsečík tak bývá ve většině případů ve všech časech měření fixován na 1 a většina úvah se pak točí kolem toho, jak nastavit váhy (a tím metriku času) u latentního průsečíku.

Kromě výchozího kódování $T(t) = \{0, 1, 2, 3 \dots j\}$ můžeme v případě měření v konstantních intervalech kódovat $T(t) = \{0, 1/j, 2/j, 3/j \dots, 1\}$, tedy tak, že první měření je kódováno 0, poslední 1 a všechna měření mezi jako zlomky času uplynulého mezi prvním a posledním měřením. Jednotkou času je tak délka trvání celé studie a takovém případě pak hodnota latentní směrnice udává změnu Y mezi prvním a posledním měřením. Takto kódovaný čas nijak nezmění shodu modelu s daty, pouze pozmění hodnoty latentní směrnice. V obou předchozích případech bylo 0 kódováno první měření. Pokud je tomu tak, pak individuální hodnota latentního průsečíku je vlastně počáteční hodnotou Y daného jednotlivce (a průměrná hodnota latentního průsečíku je průměrnou počáteční hodnotou Y v celém vzorku). Volba toho, který čas měření kódujeme 0, je tedy volbou toho, ve kterém čase bude hodnota latentního průsečíku bezprostředně interpretovatelná. Pokud bychom tedy chtěli, aby hodnota latentního průsečíku udávala hodnotu Y na konci studie, mohli bychom kódovat $T(t) = \{-j, \dots, -3, -2, -1, 0\}$. Stejně tak můžeme někdy chtít ukotvit své interpretace uprostřed časového intervalu pokrytého studií, např. $T(t) = \{-4, -3, -2, -1, 0\}$ v případě 5 měření v pravidelných intervalech. Bez ohledu na to, kam umístíme 0 je latentní tempo růstu (průměr η_1) stále stejné, jak ukazuje Tabulka II.1 – co se mění je latentní průsečík η_0 a jeho interpretace, ať již uvažujeme o jeho populačním průměru nebo individuální hodnotě.

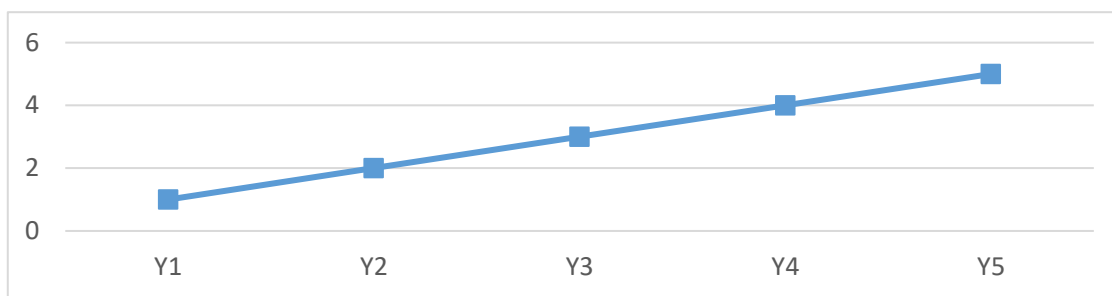
Kódování času může a nemusí odrážet skutečné jednotky času. Pokud bychom měřili například ve dvouletých intervalech, můžeme čas kódovat $T(t) = \{0, 1, 2, 3 \dots j\}$ a pamatovat si, že latentní směrnice pak udává nárůst Y za 2 roky (tj. čas uplynulý mezi dvěma časy měření, jejichž vzdálenost je kódována jako 1). Můžeme ale také jednotku času přímo zohlednit a kódovat $T(t) = \{0, 2, 4, 6 \dots j\}$. Latentní směrnice pak udává nárůst Y za jeden rok, protože vzdálenost mezi dvěma měřeními byla zakódována jako 2 (to, že jde o roky již model samozřejmě neví). Viz Tabulka II.1, specifikace S5 a S6.

Tabulka II.1. Různé způsoby kódování časových bází $T(t)$ pro lineární růst od 1 do 5 tempem 1 za jednotku času.

Fixovaný efekt latentní směrnice na manifestní (kódování $T(t)$)									
Specifikace:		S1	S2	S3	S4	S5	S6	S7	S8
η_1 na	Y1	0	0	-4	-2	0	-8	0	0
	Y2	1	0,25	-3	-1	2	-6	1	
	Y3	2	0,5	-2	0	4	-4		
	Y4	3	0,75	-1	1	6	-2	3	3
	Y5	4	1	0	2	8	0	4	4
Odhady vybraných parametrů									
	Průměr η_0	1	1	5	3	1	5	1	1
	Průměr η_1	1	4	1	1	0,5	0,5	1	1

Pozn. Všech osm různých specifikací popisuje stejnou růstovou křivku (přímku).

Obrázek II.1. Růstová přímka popsaná časovými bázemi v tabulce II.1.



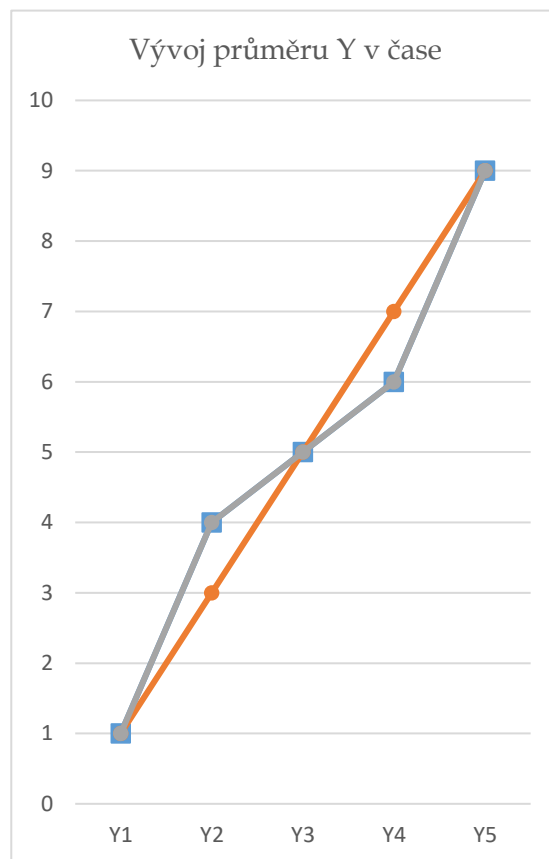
Intervaly mezi měřeními nemusí být nutně stále stejné a kódování času $T(t)$ to pak reflektuje. Pokud tedy mezi 1. a 2. měřením uplyne nějaký čas a mezi 2. a 3. měřením uplyne dvojnásobek času, pak můžeme kódovat například $T(t) = \{0, 1, 3\}$.

Zatím byly všechny časové báze pevně stanoveny, což je výhodné jak z hlediska formulace hypotézy o tvaru růstové křivky (přímky), tak z hlediska odhadu parametrů modelu; námi pevně stanovené parametry se nemusí odhadovat a informace obsažená v datech se může využít pro přesnější odhad jiných parametrů modelu. Nemusí tomu tak úplně nutně být, máme-li dostatek měření, pak je možné pevně stanovit (fixovat) pouze část hodnot času, např. první a poslední $T(t) = \{0, ?, ?, ? \dots j\}$, a ostatní hodnoty nechat odhadnout z dat. Lze si klást otázku, proč bychom něco takového chtěli dělat, když přeci většinou víme, v jakých časových rozestupech jsme data sbírali. Pokud bychom data sbírali v pravidelných intervalech a průměrná růstová křivka by vskutku byla lineární, pak by ve výše uvedeném případě byly nezafixované časové báze odhadnuty na hodnoty 1, 2 a 3. Pokud by ovšem trend nebyl lineární, pak by byly odhadnuty na hodnoty více či méně odchýlené od 1, 2, 3. Toho lze využít jednak jako test hypotézy o linearitě trendu, ale také explicitně k modelování trendů, o nichž dopředu víme, že nejsou lineární. Tyto modely s jen částečně fixovanou metrikou času se někdy nazývají *latent-base*, *level-and-shape*, nebo *latent shape* modely. Interpretace toho, jak se na odhadovaných hodnotách Y v čase t projeví hodnoty odhadnutých časových parametrů není příliš složitá. Latentní směrnice stále udává nárůst Y za časovou jednotku. Je-li časový parametr pro daný čas odhadnutý jako vyšší, než by odpovídalo uplynulé době (např. 1,5 po uplynutí jedné časové jednotky), znamená to, že v tomto konkrétním časovém intervalu byl nárůst vyšší, než by odpovídalo lineárnímu růstu (např. 1,5násobně vyšší). Obecně v daném časovém intervalu je nárůst násobkem hodnoty latentní směrnice a odhadnutého časového parametru. Tento případ je ilustrován v tabulce II.2, která ukazuje odhady časových parametrů pro růstový trend ve tvaru S.

Pomocí časových bází, které nabývají jiných hodnot, než je přesně čas/věk, v něm bylo provedeno měření, lze tedy modelovat nelineární růst pomocí modelu, který ve svých parametrech zůstává lineární (se všemi výhodami, které z toho plynou). Tato nelinearita nemusí být empiricky odhadována, jak je uvedeno výše, ale můžeme ji přímo specifikovat. Tímto způsobem například (Grimm & Ram, 2009) specifikují sigmoidní růst – časové báze jsou v modelu vypočítány z parametrů např. logistického růstu, přičemž se odhadují pouze parametry logistického růstu (jeho střed a strmost), na něž jsou časové báze pomocí modelových omezení (*constraints*) navázány.

Tabulka II.2. Časové báze pro lineární a nelineární růst

Model:		Specifikace 1 - lineární	Specifikace 2 - nelineární
Hodnota fixovaných parametrů			
Efekt latentního průsečíku na Y			
I	Y1	1	1
	Y2	1	1
	Y3	1	1
	Y4	1	1
	Y5	1	1
Efekt latentní směrnice na Y (kódování $T(t)$)			
S	Y1	0	0
	Y2	1	1,5
	Y3	2	2
	Y4	3	2,5
	Y5	4	4
Odhady vybraných parametrů			
	Průměr I	1	1
	Průměr S	2	2
Odhadnuté průměrné hodnoty manifestních (Y')			
	Y'1	1	1
	Y'2	3	4
	Y'3	5	5
	Y'4	7	6
	Y'5	9	9



Protože primárním meritorním zájmem badatelů je obvykle zjištění tvaru průměrné vývojové křivky v populaci, zůstává často poněkud v pozadí ta část modelu, která popisuje různé druhy náhodných odchylek od této křivky. To, že zůstává v pozadí se projevuje obvykle tak, že badatelé akceptují výchozí nastavení analytické procedury ve zvoleném prostředí, aniž by se zamýšleli nad tím, nakolik tato nastavení odpovídají teoretickým očekáváním. Promyšlenou součástí specifikace modelu by tedy mělo být i to, zda všechny růstové parametry mají být specifikované jako náhodné s normálním rozložením a zda tyto náhodné parametry mohou korelovat. (např. zda v modelu může být tempo růstu zčásti závislé na počáteční úrovni měřeného rysu). Zároveň je potřeba se rozhodnout na vhodné specifikaci reziduální struktury u výše uvedených modelů, která má nezanedbatelný vliv na odhad růstových parametrů modelu, shodu modelu s daty i konvergenci (Grimm & Widaman, 2010, Diallo, Morin, & Lu, 2016).

SPECIFIKACE REZIDUÁLNÍ STRUKTURY (STOCHASTICKÁ ČÁST MODELU)

Bez ohledu na to, jak je specifikována růstová (strukturní) část modelu (éta, psi), pokud modelujeme více měření, než je polynomiální stupeň křivky, nemohou růstové křivky predikovat přesně individuální naměřené hodnoty v každém čase. Také úvaha o omezené reliabilitě měření implikuje nutnost nenulových reziduálních rozptylů jednotlivých měření. Víceúrovňový regresní modelovací přístup (GCM) a přístup založený na strukturním modelování (LGC) nabízejí odlišné možnosti, jak tyto reziduální hodnoty modelovat (Grilli & Varriale, 2014).

V GCM je obvyklým výchozím předpokladem homoskedasticita a nezávislost reziduí². To je z hlediska budování modelu parametricky velmi úsporný předpoklad, protože veškerá rezidua jsou popsána jedním parametrem. Předpokládá se tedy, že odchylky od individuálních růstových křivek jsou náhodné a mají normální rozložení s rozptylem, který je v čase konstantní – homoskedastický. Také se často předpokládá, že rezidua v čase t jsou zcela nezávislá na reziduiích v čase $t-1$ a předchozích časech. To znamená, že korelace mezi opakovanými měřeními modelované charakteristiky jsou zcela vysvětlené růstovými křivkami. Reziduální hodnota je zde projevem jak nedokonalé reliability měření, tak náhodných vlivů, které vychýlily v čase t hodnotu modelované charakteristiky od modelem předpokládané hodnoty³. Vyjadřuje tím samozřejmě i nedokonalost modelu – pokud specifikujeme růstovou křivku, která neodpovídá realitě, povede to k nárůstu reziduálního rozptylu v časech, kdy se bude průměrná křivka v čase t odchylovat od průměrné hodnoty modelovaného rysu v čase t . Uvědomíme-li si tyto 3 složky reziduálního rozptylu – nereliabilitu, náhodně působící vlivy a špatnou specifikaci křivky (strukturní části modelu) – je již poněkud obtížnější věřit, že právě v našem případě lze homoskedasticitu a nezávislost reziduí předpokládat.

Vliv reliability na rezidua. Aby byl předpoklad homoskedasticity reziduí plauzibilní, je třeba předpokládat konstantní reliabilitu měření v čase, tedy invarianci měření v čase. Pokud modelujeme růstovou křivku z latentních skóru, kdy model měření je součástí modelu růstových křivek (curve-of-factors model), je tuto hypotézu možné přímo testovat. Pokud tvoříme růstový model z manifestních proměnných, je potřeba se o stabilitě reliability v čase ujistit z publikované literatury nebo doplňujících analýz. U reliability měření lze uvažovat řadu faktorů, které by mohly předpoklad konstantní reliability zpochybňovat. Je jen málo důkazů o tom, jak se na reliabilitě a validitě self-reportové škály odráží opakované vyplňování, a nakolik je faktorový model (CFA) invariantní v čase (viz např. Fried et al., 2016, Fonseca-Pedrero et al., 2010). Pokud tedy nemůžeme předpokládat konstantní reliabilitu napříč měřeními, je na místě uvažovat různé heteroskedastické struktury. To může například znamenat volně odhadovaná, avšak stále nezávislá rezidua v každém čase měření. To je výchozí reziduální struktura v LGC modelu v Mplus. Zcela volná rezidua, která se můžou lišit v každém čase měření, jsou možná extrémem, pro který existuje opodstatnění spíše v případě

² V případě opakovaných měření v pevně daných časech je i v rámci multilevel lineárního modelu možné specifikovat i jiné struktury než homoskedastická nezávislá rezidua, ale namísto úplné volnosti je potřeba si obvykle vybrat z palety nabízených obvyklých kovariančních struktur (matic).

³ Pokud by vychylující vlivy nebyly náhodné a samy také měly nějakou stabilitu v čase, pak jimi způsobené odchylky sledované charakteristiky od růstové křivky budou v čase korelovat – a rezidua tak přestanou být nezávislá.

dat z experimentálních designů. Pro každý čas měření je potřeba jeden parametr navíc, a to může být při větším počtu měření zbytečná zátěž pro odhad důležitějších parametrů modelu. McArdle & Nesselroade (2014) nedoporučují se od homoskedasticity vzdalovat, byť jen z důvodu srovnatelnosti výsledků napříč různými typy modelů (s. 103). Přesto je to výchozí nastavení v některých programech pro strukturní modelování, zejména v Mplus. Grimm a Widaman (2010) nabízejí dvě specifikace reziduálních struktur, které umožňují postupně se měnící rezidua v souladu s teoretickým předpokladem. První z nich je „konstantní reliabilita“, čímž se míní předpoklad, že s postupem času se nemění podíl rozptylu závislé vysvětlený růstovým modelem. Tento podíl nazývá růstová reliabilita, *growth curve reliability* podle Hertzog, Ghisletta, Lindenberger, & von Oertzen (2006). Druhou je lineárně se měnící (rostoucí, klesající) růstová reliabilita, což je struktura odpovídající předpokladu, že podíl rozptylu vysvětlený růstovým modelem se v čase postupně mění, ať již z důvodů měnící se reliability měření, nebo z důvodu nesouladu mezi specifikovanou růstovou křivkou. V obou případech jde při větším počtu měření o parametricky úspornější specifikace reziduální struktury.

Vedle rozptylů reziduí je dále nutné uvažovat také kovariance reziduí. V předchozím textu jsme zatím uvažovali nezávislá, nekorelovaná rezidua. To odpovídá předpokladu, že jakákoli korelace mezi opakovanými měřeními je dána pouze existencí v čase se měnícího se rysu modelovaného růstovou částí modelu (podmíněná nezávislost). Rozptyl reziduí se pak skládá z náhodná chyb měření a nesystematických (náhodných) situačních vlivů na modelovaný rys. Tyto dvě složky obvykle v modelu nejsou nerozlišeny. Někdy je ale na místě předpokládat, že jednotlivá měření v čase nejsou podmíněně nezávislá. Lze tedy uvažovat třeba autoregresivní struktury (AR, MA, Toeplitz, apod.), kdy dřívější měření ovlivňuje hodnotu následujícího měření nad rámec stability rysu v čase, přičemž korelace mezi rezidui se s rostoucí vzdáleností mezi časy měření snižuje.

Konečně, nelze zcela zapomenout ani na možnost, že na modelovaný konstrukt dlouhodobě působí nějaký neměřený korelát, který způsobuje, že rezidua korelují výrazně nad rámec toho, co by implikoval růstový model. V nejhorším případě může být tento korelát dokonce nežádoucí složkou sumačního manifestního skóru, jehož vývoj modelujeme - vždyť čistě unidimenzionální měřítka rozhodně nejsou normou.

Jak se tedy ohledně volby reziduální struktury v praxi rozhodovat? Hedeker & Gibbons (2006) doporučují porovnat různé reziduální struktury proti zcela nestrukturované reziduální struktuře pomocí likelihood ratio testu. Zajímavé je, podle těchto autorů je vhodné to dělat na modelu, do kterého jsou zahrnuty všechny potenciální kovariáty (s. 129).

2. Provedení analýzy latentních tříd růstových křivek

Cílem analýzy latentních růstových tříd je modelování heterogenity růstových trajektorií pomocí dvou nebo více tříd, v nichž jsou růstové parametry konstantní. Oproti LGC modelu tedy musíme do LCGA modelu zařadit nominální proměnnou (obvykle pojmenovanou C), která bude reprezentovat příslušnost jednotlivce k latentní třídě. Tato proměnná reprezentující latentní třídy vlastně predikuje růstové parametry (průsečík, směrnici apod.). Model se odhaduje vždy pro určitý specifikovaný počet latentních tříd. Obvykle začínáme modelem se dvěma třídami a pak postupně počet tříd navyšujeme, dokud je řešení smysluplné, nebo dokud model konverguje ke spolehlivým a nezkresleným odhadům parametrů. Tímto způsobem získáme model, v němž bude mít každá třída svůj vlastní odhad růstových parametrů.

I když se může zdát, že předpoklad, že uvnitř tříd v LCGA modelech je nulová variabilita růstových trajektorií, je z teoretické perspektivy ve většině situací obtížně obhajitelný, často prakticky neměříme tak přesně a vývoj je tak složitý, že je výhodné přesunout tuto heterogenitu do reziduální části modelu. Existuje tak řada studií, které prezentují právě třídy z LCGA modelů (např. Dowsey, Smith, & Choong, 2015, či De Vos, Runhaar, Verkleij, Van Middelkoop, & Bierma-Zeinstra, 2014)

Latentní třída však nemusí ovlivňovat jen parametry růstu, ale v principu vlastně všechny parametry LGC modelu. V extrému bychom mohli hledat pro každou latentní třídu zcela samostatný LGC model. Rozhodnout se musíme zejména o tom, jak má být v modelu latentní třídou ovlivňována reziduální struktura. Obvykle nejprve specifikujeme model s reziduální strukturou invariantní napříč třídami. To znamená, že třídy se liší pouze růstovou trajektorií a reziduální variabilita kolem růstové trajektorie je ve všech třídách stejná. Za řady okolností však teorie predikuje že určitá podoba růstu může být více stabilní než jiná a má tedy smysl specifikovat model tak, že rezidua se mohou napříč třídami lišit. Rámec strukturálního modelování nám zde dává obrovskou flexibilitu, která je však z hlediska mixture modelů poněkud zrádná. Můžeme specifikovat model tak, že se reziduální struktury liší napříč latentními třídami, ale i tak, že se třeba jen jedna třída ve své reziduální struktuře liší od ostatních. V závislosti na tom, jak složitou reziduální strukturu jsme specifikovali ve výchozím LGC modelu (konstantní vs. v čase proměnlivá rezidua, korelace reziduí atd.), pak následuje odpovídající násobný nárůst počtu parametrů. Zároveň, čím složitější reziduální struktura s více parametry, tím spíše nalezneme teoretické důvody, proč by alespoň některé parametry neměly být invariantní napříč třídami. V této etapě analýzy lze snadno model „přeparametrizovat“ a dostat se do potíží s identifikací a konvergencí a s nimi spojenými časovými problémy. Obvykle je tedy potřeba hledat kompromis mezi detailností reziduální struktury a specifikováním rozdílů v reziduální struktuře napříč latentními třídami. Toto hledání kompromisu je obtížné také proto, že specifikace (in-)variantnosti reziduální struktury má často velký dopad na to, jaké třídy budou „extrahovány“ a také na ukazatele shody modelu s daty.

Příklady specifikace různě odlišných latentních tříd v Mplus, jak byly použity v analýzách v následujících kapitolách, jsou uvedeny v příloze 1.

3. Specifikace růstového mixture modelu

Zatímco LCGA modely předpokládají, že růstové parametry jsou uvnitř tříd konstantní, růstové mixture modely předpokládají, že růstové parametry uvnitř tříd⁴ mají normální rozložení s odhadovaným rozptylem a možnými kovariancemi. Pomocí nenulových rozptylů růstových parametrů tak jedna třída popíše širší skupinu jedinců. Růstová trajektorie reprezentující třídu je pak *průměrnou* růstovou trajektorií v této třídě.

S nenulovými rozptyly růstových parametrů přibývá další volba pro specifikaci modelu. Je nutné se rozhodnout, zda předpokládat, že se jednotlivé třídy budou lišit tvarem průměrné růstové křivky i variabilitou růstových křivek uvnitř třídy nebo zda se budou lišit pouze tvarem průměrné růstové křivky. Wickrama et al., (2016) pro tyto dvě možnosti užívají termíny *class-variant* a *class-invariant* GMM (GMM-CV, GMM-IV).

⁴ Zde bych vlastně již neměl používat termín *třída*, ale *komponenta* směsi. Zatímco třídy reprezentují konstanty, komponenty směsi jsou (normální) distribuce. Ve statistické praxi se však běžněji používá termín *třída* i v růstových mixture modelech.

Protože i při nejjednodušším lineárním růstu znamená GMM-CV tři odhadované parametry navíc pro každou třídu, je zřejmé, že toto rozhodnutí nelze dělat lehkomyšlně a bez přihlídnutí k teorii. Model se snadno stane příliš složitým, což se projeví nestabilitou parametrů a problémy s konvergencí. Hipp & Bauer (2006) uvádějí, že napříč třídami invariantní rozptyly (růstových parametrů nebo reziduální) výrazně snižují výskyt problémů s lokálními řešeními, a to jak z hlediska konvergence, tak z hlediska pravdivosti, respektive replikovatelnosti parametrů, které se mohou lišit napříč třídami.

Zde by již bylo na místě zrekapitulovat volby o specifikaci parametrů napříč třídami. *Rezidua*: ať již se rozhodneme pro jakoukoli reziduální strukturu, je potřeba se rozhodnout, zda bude napříč třídami shodná (což je výchozí doporučovaná podoba), nebo odlišná. Je mnoho možných podob odlišnosti reziduálních struktur napříč třídami. Mohou se lišit pouze velikosti reziduálních rozptylů, ale i jejich struktura. Mohou se mezi sebou lišit všechny skupiny, nebo jen jedna od ostatních. Možností je tolik, že je nutné se řídit teoretickými předpoklady. *Specifikace tvaru růstové křivky*: Růstová křivka může mít napříč třídami stejný tvar, ale může se i lišit. V závislosti na komplexitě specifikace lze libovolné parametry napříč třídami fixovat. *Rozptyly růstových parametrů*: Opět mohou být stejné, mohou se lišit, a totéž platí pro jejich kovariance.

Vedle voleb ohledně specifikace parametrů růstových křivek a možných rozdílů mezi třídami stojí před badatelem ještě další důležité rozhodnutí, a to zda určovat nejpravděpodobnější počet tříd v populaci (vzorku) na nepodmíněném modelu, kde jsou indikátory tříd pouze latentní parametry růstových křivek, nebo zda již od počátku řešit otázku počtu tříd modelem obsahujícím kovariáty, související proměnné, které by se podle teorie měly podílet na utváření tříd (Tofighi & Enders, 2006). V současnosti se na poli GMM modelů lze setkat se dvěma protichůdnými názory na tuto otázku. Někteří autoři na základě simulačních studií doporučují kovariáty do enumeračního modelu zahrnovat (např. Li & Hser, 2011), protože dobře zahrnuté kovariáty zvyšují pravděpodobnost správného určení počtu tříd a snižují problémy s identifikací parametrů modelu a konvergencí odhadu (Muthén, 2004). Jiní, opět na základě simulačních studií, (např. Nylund-Gibson & Masyn, 2016) považují takovou volbu za příliš nebezpečnou, protože v případě nesprávné specifikace role kovariátu v modelu dojde naopak k výraznému zhoršení kvality extrakce tříd. Vzhledem k omezené teorii a široké paletě možných rolí kovariátu je nesprávná specifikace bohužel poměrně pravděpodobná. Více o zakomponování kovariátů v kapitole II.6.

4. Řešení problémů s odhadem parametrů modelů

Při odhadu parametrů mixture modelů dochází velmi často k situaci, kdy estimační algoritmus konverguje k nějakému lokálnímu maximu likelihoodové funkce (Muthén & Shedden, 1999). Hipp & Bauer (2006) uvádí, že tento problém si lze intuitivně představit tak, že některé kombinace odhadů parametrů modelu dobře vystihují dobře část dat, ale nevystihují nejlépe data jako celek. Parametry tzv. lokálního řešení se mohou od parametrů globálního řešení podstatně lišit, a to nejen z hlediska své věcné interpretace, ale i z hlediska replikovatelnosti a v konečném důsledku „pravdivosti“ řešení, pokud je cílem identifikovat reálně existující skryté subpopulace.

Rozsah problému s lokálními řešeními je větší, než si obvykle představujeme. Nejde o pouze o zvýšenou pravděpodobnost, že parametry modelu, který úspěšně konvergoval, mohou reprezentovat jen nějaké lokální řešení. Hipp & Bauer (2006) reanalyzovali analýzy publikované Muthénovými (Muthén & Muthén, 2000) a detailně prozkoumali celou množinu

možných hodnot všech parametrů (*parameter space*), tedy spustili odhad modelu 999krát se startovacími hodnotami pokrývajícími celé rozpětí možných hodnot odhadovaných parametrů. U LCGA modelu se 6 třídami našli 30 různých jedinečných řešení a tento počet narostl u 9 tříd až na 81 jedinečných řešení. U GMM modelu bylo již u čtyř tříd 11 jedinečných řešení, ale počet jedinečných řešení již s rostoucím počtem tříd nerostl tak rychle jako u analogického LCGA. To znamená, že cílem odhadu není dosáhnout konvergence, ale dosáhnout konvergence mnohokrát, tolikrát, abychom mohli s rozumnou jistotou vybrat řešení, které konvergovalo k nejmenšímu likelihoodu – pravděpodobně globálnímu minimu, respektive řešení.

Tento problém se obvykle řeší opakovaným spouštěním estimačního algoritmu s různými, obvykle náhodně generovanými startovacími hodnotami pro jednotlivé parametry. To, že s různými startovacími hodnotami dosáhne algoritmus různých maximálních hodnot je známkou problému s lokálními minimy. Pokud se podaří s různými náhodnými startovacími hodnotami opakovaně dosáhnout stejného minima likelihoodové funkce, je naděje, že by mohlo jít o hledané globální minimum. Replikace minimálního likelihoodu s různými startovacími hodnotami samozřejmě není postačující (a vlastně ani nutnou) podmínkou pro jistotu, že minimum je globální, ale šance zvyšuje.

Statistické programy umožňující odhad parametrů mixture modelů často defaultně nabízí opakované odhadování s různými startovacími hodnotami. S ohledem na výpočetní náročnost mixture modelů je počet opakovaných odhadů mixture modelu s náhodnými startovacími hodnotami spíše nižší, než je pro rozumnou míru jistoty potřeba. Například Mplus má výchozí počty opakování v řádu desítek, ale v každém výstupu z analýzy nabádá uživatele k nejméně zdvojnásobení tohoto počtu, a to bez ohledu na to, zda uživatel již počet opakovaných odhadů navýšil. Obecně je i u poměrně jednoduchých longitudinálních modelů potřeba počítat spíše se stovkami a až tisíci opakovaných odhadů s náhodnými startovacími hodnotami parametrů.

Může nastat i „spurious“ globální minimum – tedy model se skokově nižším likelihoodem, který je ale obtížně replikovatelný a patrně je pouze artefaktem konkrétních dat (Hipp & Bauer, 2006). Tueller & Lubke (2010) navíc doporučují prozkoumat i odhady, které vedly k druhému/třetímu nejmenšímu minimu likelihoodové funkce, a porovnat jejich parametry s parametry odpovídající nejvyššímu dosaženému minimu, abychom se ujistili o stabilitě řešení.

5. Volba optimálního nepodmíněného modelu s třídami

Ve všech modelech směsí je jednou ze základních otázek identifikace počtu latentních tříd v populaci (McLachlan & Peel, 2000). Předpokládá se, že v populaci existuje k tříd, z nichž výběrem s nějakou výběrovou chybou (a také s nějakou chybou měření, která je explicitněji modelována až v modelech vyššího řádu) vzniknou pozorovaná data, a naším úkolem je zpětně na počet tříd v populaci usuzovat. Ani na jednoduchých unidimenzionálních datech nemusí být triviální takové rozhodnutí učinit. Pozorovaná směs k komponent může být empiricky nerozlišitelná od směsí s menším či větším počtem komponent, v závislosti na tom, jak velké a odlišné komponenty mohou být. Prakticky se tato otázka redukuje na otázku po minimálním počtu tříd (komponent), které dobře pasují na data (McLachlan & Peel, 2000). Nejčastější metrikou shody modelu s daty je věrohodnost, tedy podmíněná pravděpodobnost pozorovaných dat za konkrétních hodnot parametrů modelu $P(Y|\theta)$.

Ani takto redukována otázka nemá jednoduché řešení a matematikové navrhli velké množství ukazatelů, které je možné při rozhodování o počtu latentních tříd. McLachlan & Peel (2000) ve své základní práci referují o třech hlavních kategoriích ukazatelů. První jsou ukazatele založené na informačních kritériích, tedy hodnotě věrohodnosti penalizované za množství volných parametrů v modelu. Druhé jsou LRT testy používající principy NHST pro test hypotézy, že fit modelu s k třídami je lepší než fit modelu s $k-1$ třídami. Poslední kategorií jsou ukazatele založené na entropii, tedy jasnosti kategorizace do tříd, separaci tříd.

I když je nečastějším postupem enumerace počtu latentních tříd vzájemné porovnávání modelů s postupně rostoucím počtem tříd, není to jediný a podle Liu & Hancock (2014) ani optimální postup. Liu & Hancock (2014) testují postup navržený Bauerem a Curranem (2004), kdy se počet latentních tříd stanovuje na zcela volném (unrestricted) modelu, a až po stanovení nejpravděpodobnějšího počtu latentních tříd se pro tento počet tříd specifikuje GMM model. Dochází k tomu, že tradičním postupem patrně často dochází k nadhodnocené představě o počtu latentních tříd v populaci, zároveň však uvádí, že odhady volném modelu naráží i na velkých vzorcích na velké konvergenční problémy, jejichž komplikované řešení snižuje využitelnost tohoto postupu.

INFORMAČNÍ KRITÉRIA

Informační kritéria vyjadřují shodu modelu s daty prostřednictvím hodnoty likelihoodové funkce. Čím více je model flexibilní, čím má více volných parametrů, tím větší je prostor pro shodu s daty. A modely s latentními třídami mohou mít mnoho parametrů. Proto je potřeba flexibilnější modely penalizovat, aby bylo možné dosáhnout nějaké míry parsimonie. Obvykle se pracuje se záporným dvojnásobkem minima věrohodností funkce ($-2LL$), která vyjadřuje míru diskrepance mezi pozorovanými daty a modelem (jsou tedy škálovány tak, že nižší hodnoty jsou lepší) ke kterému jsou přičítány různé penalizace. Například hojně používané Akaikeho informační kritérium (AIC, Akaike, 1974), které přidává k $-2LL$ dvojnásobek počtu volných parametrů v modelu (p):

$$AIC = -2LL + 2p$$

S ním je příbuzný konzistentní AIC (CAIC, Bozdogan, 1987), který počet parametrů násobí o jednu zvětšeným logaritmem velikosti vzorku.

$$CAIC = -2LL + p(\log(N)+1)$$

V kontextu mixture modelů se často doporučuje Schwarzovo Bayesovské informační kritérium (BIC, Schwarz, 1978), které přidává k $-2LL$ počet volných parametrů v modelu vynásobený logaritmem velikosti vzorku.

$$BIC = -2LL + p \log(N)$$

Hedeker & Gibbons (2006) uvádí, že u BIC není tak úplně jasné, zda se zde N míní počet zkoumaných osob, nebo počet měření. S odkazem na Rafteryho (1995) a poukázáním na výstup SAS uvádí, že konsenzuálně se spíše považuje za N počet zkoumaných osob. Na druhou stranu SPSS v MIXED proceduře používá jako N počet měření (platných řádků v dlouhých datech) a je-li růstový model specifikován jako multilevel (*mixed random model terminologií Heddeker a Gibbonse*), tak i M plus používá počet měření. V kontextu růstových modelů nejde o triviální otázku kvůli chybějícím datům. Hedeker & Gibbons (2006) dále uvádí, že penalizace BIC je poměrně přísná v případě, že posuzujeme modely lišící se pouze specifikací kovarianční struktury reziduí, což vede k preferování modelů s příliš jednoduchou kovarianční strukturou. V takových případech doporučují méně penalizující AIC.

Modifikací BIC je Sample size adjusted BIC (SABIC či SSABIC, Sclove, 1987), který také penalizuje za velikost vzorku, ale mírněji, než BIC.

$$\text{SABIC} = -2LL + p \log((N+2)/24)$$

U mixture modelů vychází v řadě simulačních studií (Tofighi & Enders, 2006, Yang, 2006, Henson, Reise, & Kim, 2007, Nylund, Asparouhov, & Muthén, 2007) jako ukazatel s nejvyšší silou SABIC, popř. BIC. Nověji Tein, Coxe, & Cham (2013) preferují BIC, protože na malých vzorcích (<250) ztrácel SABIC na síle. Napříč simulačními studiemi panuje shoda v tom, že nejhůře indikuje počet tříd AIC.

Informační kritéria vedou ke správné volbě počtu latentních tříd za splnění podmínky normality rozložení růstových parametrů uvnitř tříd. Při výraznějších odklonech od normality se zvýšenou pravděpodobností dochází k nadhodnocení počtu tříd a k extrahování nepravých (spurious) tříd (Guerra-Peñ, Steinley, & Guerra-Peña, 2016). Podle simulací citovaných autorů k tomu dochází i tehdy, kdy jsou non-normálně rozložené časově invariantní kovariáty přítomné v modelu při rozhodování o počtu tříd.

LIKELIHOOD RATIO TEST

Tradiční technikou porovnávání modelu je likelihood ratio test (LRT) založený na tom, že podíl likelihoodů dvou modelů realizovaný jako rozdíl jejich logaritmů ($-2LL$) má za určitých podmínek známé teoretické rozložení, jmenovitě rozložení chí kvadrát s počtem stupňů volnosti rovným rozdílu v počtu volných parametrů mezi srovnávanými modely.

$$LR = -2(LL_{\text{redukovaný}} - LL_{\text{plný}}).$$

Jednou z těchto podmínek je to, aby model s nižším počtem volných parametrů byl vnořený do modelu s vyšším počtem volných parametrů, tedy aby restriktivnější model mohl vzniknout pouhým zafixováním jednoho či více parametrů. Další podmínkou regularity je podle (Krauter, Muthén, 2008) to, aby parametry nenabývaly často hodnot blízkých mezím svých možných hodnot (*boundary condition*). Ovšem pravděpodobnost členství ve třídě je parametr nabývající často hodnot blízko svých mezí (0; 1). Dokonce ve vnořených modelech je pravděpodobnost vynechané třídy nastavena na 0. Proto nemá rozdíl dvou $-2LL$ chíkvadrát rozložení (Krauter, Muthén, 2008). Tento problém lze řešit korekcí p-hodnoty či referenční pravděpodobnostní distribuce, nebo přímo resamplingovým stanovením výběrové distribuce $-2LL$ pro restriktivnější model.

Lo, Mendel, Rubin (2001) se vydali v návaznosti na Vuonga (1989) tou první cestou a rozdělení $-2LL$ $k-1$ modelu specifikovali jako váženou sumu nezávislých chí-kvadrát rozložení s jedním stupněm volnosti. Test LMR (Lo-Mendel-Rubin) je implementován v Mplus (příkaz TECH11; v sekci OUTPUT:). Funguje tak, že v rámci odhadu modelu s k třídami je dodatečně ještě odhadnut model s $k-1$ třídami (H_0) a s použitím patřičného rozložení je stanovena pravděpodobnost rozdílu $-2LL$ mezi k a $k-1$ modely při platnosti H_0 . Nízká p-hodnota zamítá model s $k-1$ třídami ve prospěch modelu s k třídami. Pro účel odhadu $k-1$ modelu je v Mplus volba K-1STARTS v sekci ANALYSIS, která umožňuje nastavit množství náhodných startovacích hodnot a finálních optimalizací tak, aby ani odhad $k-1$ model neuvízl na nějakém lokálním minimu. U jednoduchých LCA/LPA modelů to nebývá až takovým problémem, ale u GMM lze očekávat problémy s lokálními minimy a konvergencí i u modelů s teoreticky správným počtem tříd (ve smyslu H_0).

Protože původní verze testu nepodávala dobré výkony, navrhli k ní Lo, Mendel a Rubin ještě ad hoc korekci. Podle Tofighiho a Enderse (2006) však nemá korekce velký dopad. Mplus

udává jak originální test, tak jeho korigovanou verzi. Pokud není uvedeno jinak, mini se LMR testem jeho korigovaná verze. LMR předpokládá multivariační normalitu závislé (podmíněnou na případných kovariátech) a v případě porušení tohoto předpokladu má tendenci preferovat modely s příliš mnoha třídami (Muthén, 2003). Podle Nylundové, Asparouhova a Muthéna (2007) je LMR test poměrně přesným ukazatelem počtu tříd u GMM, zvláště pak na vzorcích vyšších než 500 jednotek (>90 % úspěšnost). Ještě lepší úspěšnost měl však v jejich studii bootstrapový LRT.

Bootstrapový LRT (BLRT) byl navržen specificky pro účel porovnávání modelů směsí (McLachlan & Peel, 2000), který pomocí bootstrapování odhadne rozložení dvojnásobku rozdílu likelihoodů srovnávaných modelů, kterou použije k výpočtu p-hodnoty. Interpretace BLRT testu je stejná jako u LRT testu porovnávajícího model s $k-1$ třídami s modelem s k třídami. Tento test je vedle BIC/SABIC považován za nejlepší statistické kritérium toho, zda další třída již není overfittingem. Bohužel je výpočetně velmi náročný, a proto je důležité zvládnout postup jeho využití a nastavení. Postup výpočtu BLRT testu je podle (Asparouhov & Muthén, 2012) následující:

- Při odhadu modelu s k třídami se odhadne také model s $k-1$ třídami a spočítá se dvojnásobek rozdílu log-likelihoodů. V situaci, kdy stanovujeme specifické restriktce na parametry modelu v jednotlivých třídách, je důležité vědět, že model s $k-1$ třídami je v rámci implementace BLRT testu v Mplus specifikován tak, že se vypustí první třída, jak je specifikovaná v syntaxu (Asparouhov, Muthén, 2012). Jde-li v našem konkrétním případě o třídu se zvláštními restrikcemi parametrů, bude se model s $k-1$ třídami věcně lišit od modelu s k třídami, tedy nikoli jen počtem tříd.
- Parametry modelu s $k-1$ třídami se použijí k vygenerování dat odpovídajících $k-1$ modelu. Na těchto datech se odhadne jako model s $k-1$ třídami (který by měl pasovat velmi dobře), tak model s k třídami. Opět se spočítá dvojnásobek rozdílu jejich log-likelihoodů. Tento krok se mnohokrát zopakuje, čímž získáme odhad rozložení $-2LL$ při platnosti nulové hypotézy, že pravdivý počet tříd je $k-1$.
- Rozdíl $-2LL$ mezi modelem s k třídami a modelem s $k-1$ třídami zjištěný v prvním kroku na empirických datech je pak porovnán s bootstrapovým rozložením $-2LL$ a stanovena p-hodnota. Dozvídáme se tedy, jak pravděpodobné je, že by nám $-2LL$ mezi k a $k-1$ modelem vyšel tak velký, kdyby byl pravdivý model s $k-1$ třídami.

Prostřední krok této procedury může být v závislosti na počtu generovaných vzorků výpočetně velmi náročný. Samotný odhad modelů směsí je relativně náročný a trvá podle komplexnosti modelu řádově minuty či desítky minut. Pakliže se má model na generovaných datech odhadovat řádově stokrát, čas výpočtu BLRT se snadno dostane do řádu hodin.

Pro efektivní výpočet BLRT testu je vhodné se držet doporučení Asparouhova a Muthéna (Asparouhov & Muthén, 2012) a nejprve nechat odhadnout model s k třídami s takovými nastaveními, aby bylo poměrně jisté, že bylo dosaženo globálního minima likelihoodové funkce. To se týká zejména nastavení počtu náhodných startovacích hodnot a finálních optimalizací (volba STARTS v sekci ANALYSIS). Pro generování každé náhodné sady startovacích hodnot je využito semínko (*seed*), které je uvedeno ve výstupu Mplus. Semínko je následně možné vložit zpět do syntaxu (OPTSEED=, STARTS=0), čímž je dosaženo toho, že model se s tímto nastavením odhadne vždy zcela stejně napoprvé, bez opakovaného odhadování s různými startovacími hodnotami. Až s nastaveným OPTSEED má smysl v sekci OPUTPUT požádat o TECH14, což je žádost o výpočet BLRT testu.

Protože problémy s lokálními minimy a konvergencí se týkají i odhadování modelů na datech vygenerovaných v rámci BLRT testu, je dobré vhodně nastavit počet náhodných sad startovacích hodnot a finálních optimalizací i pro tyto modely. K tomu slouží volby LRTSTARTS a K-1STARTS. LRTSTARTS udává počet startovacích hodnot a optimalizací pro $k-1$ model; druhá dvojice čísel udává totéž pro k model. Výchozí hodnoty pro $k-1$ model jsou „0 0“, protože se odhaduje zaručeně správný počet tříd (pro které jsme generovali data), a tak se neočekávají problémy s lokálními minimy. Přesto i zde může u složitějších modelů docházet k problémům. Nastavení na 20 4 příliš nezdrží a umožní Mplus si uvědomit a ve výstupu uvést, že se na některých vygenerovaných datech nepodařilo u $k-1$ modelu replikovat hodnotu minimálního LL, což značí problémy s lokálními minimy. Pro k model je podle mých zkušeností dobré nastavit vyšší hodnoty, než s jakými byl odhadnut model na našich datech. Je to proto, protože zde skutečně odhadujeme špatný, o jednu třídu vyšší, počet tříd. To eskaluje problém s lokálními minimy.

K-1STARTS stejně jako u LMR testu nastavuje počet startovacích hodnot a optimalizací pro $k-1$ model odhadovaný na pozorovaných datech (tedy ne těch generovaných).

Dalším zásadním rozhodnutím je počet vygenerovaných vzorků. LRTBOOTSTRAP nastavuje počet bootstrapových vzorků, resp. vzorků generovaných z parametrů $k-1$ modelu. Zpočátku je výhodné ponechat tuto hodnotu nenastavenou. Mplus pak používá *sequential stopping rule* (viz appendix Nylund et al., 2007), kdy generuje větší počet vzorků pouze tehdy, když se p -hodnota pohybuje okolo 0,05. Pokud je velmi blízká nule, nebo výrazně vyšší, další vzorky se negenerují. Ručně se LRTBOOTSTRAP obvykle nastavuje na hodnoty mezi 100 a 500. Tekle, Gudicha, & Vermunt (2016) používají ve své studii statistické síly BLRT testu 500 bootstrapových vzorků, což je patrně minimem, za které už není účelné jít.

Pro kontrolu korektního nastavení a běhu BLRT testu je vhodné zkontrolovat, zda hodnoty log-likelihoodu uváděné BLRT testem pro k a $k-1$ model jsou totožné s těmi, které jsme v předchozích krocích analýzy zjistili při samostatném odhadu modelu s $k-1$ a modelu s k třídami.

Tofighi & Enders (2006) řeší problém pořadí tříd, protože v BLRT testu se vymazává první třída. Pokud jsou všechny třídy specifikovány stejně, na pořadí nezáleží. Pokud ale máme některé třídy specifikované odlišně, může pořadí ovlivnit výsledky testu. V takovém případě je dobré vyhnout se specifikování první třídy jako té odlišné. Lze také uvažovat tak, že enumeraci tříd provedeme na modelu se stejně specifikovanými třídami a až potom je budeme ladit. To se však nejeví jako ideální přístup, protože specifikace tříd má vliv na enumeraci.

UKAZATELE ZALOŽENÉ NA ÚSPĚŠNOSTI KLASIFIKACE

Spolu se stanovením počtu latentních tříd v populaci reprezentované vzorkem, který analyzujeme, je jedním z ústředních cílů mixture modelů klasifikace jednotlivců do latentních tříd. Kvalita, úspěšnost, jistota klasifikace jsou tak dalším důležitým kritériem pro posuzování kvality modelu latentních tříd či mixture modelu.

Úspěšnost klasifikace se odvíjí od posteriorních pravděpodobností náležení do latentní třídy $P(c=k | U_i)$, kde c je kategorická proměnná reprezentující latentní třídu, k je číslo latentní třídy a U_i je individuální vektor hodnot indikátorů latentních tříd (tedy v našem případě parametry růstové křivky daného jednotlivce.

$$P_i(c = k | U_i) = P_{ik}$$

Jde tedy o pravděpodobnost, s jakou jednotlivec náleží do latentní třídy, pokud má jeho individuální růstová křivka parametry, které má. Tyto pravděpodobnosti jsou vypočítány z rozdílů rozložení indikátorů napříč latentními třídami. Pokud bychom uvažovali lineární LGC model, kde $U_i = \{I, S\}$ (I = latentní průsečík a S = latentní směrnice),

$$P(c = k | \{I_i; S_i\}) = \frac{P(\{I_i; S_i\} | c = k) P(c = k)}{P(\{I_i; S_i\})}$$

Alternativně bychom místo $U_i = \{I, S\}$ mohli použít $U_i = \{Y_1, Y_2, \dots, Y_j\}$, tedy manifestní hodnoty modelované proměnné naměřené v jednotlivých časech.

Pro každého jednotlivece vyplývá z modelu posteriorní pravděpodobnost pro každou latentní třídu, tj. pro model se třemi latentními třídami jsou to tři posteriorní pravděpodobnosti $P(c=1 | U_i)$, $P(c=2 | U_i)$ a $P(c=3 | U_i)$. Součet posteriorních pravděpodobností pro každého jednotlivece činí 1. Každý jednotlivec je tedy členem některé z latentních tříd.

$$\sum_{k=1}^K P_i(c = k | U_i) = \sum_{k=1}^K P_{ik} = 1$$

Z perspektivy úspěšnosti, či jistoty klasifikace jsou preferovány modely, pro které platí, že každý jednotlivec (jednotka) má vysokou posteriorní pravděpodobnost náležení do jedné třídy a velmi nízké posteriorní pravděpodobnosti náležení do všech ostatních tříd. Jinými slovy, preferujeme modely, které se blíží implicitně ideálnímu výsledku klasifikace, kdy víme, do které latentní třídy každý jednotlivec patří (s vysokou pravděpodobností). To není samozřejmé; běžně se lze setkat s modely, kdy má podstatná část vzorku posteriorní pravděpodobnosti podobné pro více než jednu třídu, tedy blízké apriorním (ty jsou rovny $1/k$), z čehož je zřejmé, že model na základě individuální růstové křivky neumí daného jednotlivece kategorizovat.

Jednotlivec je nakonec modelem klasifikován do třídy, pro kterou má nejvyšší posteriorní pravděpodobnost (most-likely class, highest-probability class). Udržuje se zde distinkce mezi latentní třídou, jíž je jednotlivec skutečně členem (neznámý stav), a třídou, do které byl jedinec na základě posteriorních pravděpodobností zařazen.

Pro účely hodnocení modelu není praktické posuzovat posteriorní pravděpodobnosti jednotlivých účastníků výzkumu. Namísto procházení posteriorních pravděpodobností pro každého jednotlivece ve vzorku tuto charakteristiku modelu obvykle zachycujeme klasifikačními tabulkami, které pro skupiny jednotlivců klasifikovaných do jednotlivých tříd uvádí buď průměrné posteriorní pravděpodobnosti pro jednotlivé třídy (tj. odpověď na otázku „Jaká je *průměrná* pravděpodobnost náležení do tříd (populačně) 1, 2, ..., k u jednotlivců, kteří byli klasifikováni do třídy 1, 2, ..., k?“) nebo podmíněnou pravděpodobnost náležení do populační třídy 1, 2, ..., k, pokud byl jedinec klasifikován do třídy 1, 2, ..., k. Žádoucí je, aby v klasifikačních tabulkách byly diagonální hodnoty blízké 1 a hodnoty mimo diagonálu blízké 0.

Interpretace průměrných posteriorních pravděpodobností je relativně přímočará. Jde o průměry posteriorních pravděpodobností pro danou třídu podle toho, v jaké třídě má jedinec nejvyšší posteriorní pravděpodobnost $M(P_{ik} | \max k)$. Příklad výstupu Mplus v tabulce II.3 tedy uvádí, že průměrná posteriorní pravděpodobnost náležení do 1. třídy jednotlivců, kteří mají nejvyšší posteriorní pravděpodobnost pro 1. třídu $M(P_{11} | \max k = 1)$ je 0,94. Průměr posteriorních pravděpodobností členství v 2. latentní třídě týchž jednotlivců je 0,06, což je $1 - 0,94$. Podobně průměrná posteriorní pravděpodobnost náležení do 2. třídy jednotlivců, kteří mají nejvyšší posteriorní pravděpodobnost pro 2. třídu $M(P_{12} | \max k = 2)$ je 0,85. Problémem

průměrných posteriorních pravděpodobností podle nejpravděpodobnější latentní třídy je to, že rozložení průměrovaných pravděpodobností je obvykle vysoce zešíkmené.

Tabulka II.3. Průměrné posteriorní pravděpodobnosti ve výstupu Mplus.

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)		
	1	2
1	0.940	0.060
2	0.147	0.853

Průměrné posteriorní pravděpodobnosti podle tříd lze vlastně vnímat jako podmíněné pravděpodobnosti členství v latentní třídě, pokud je jedinec modelem klasifikován do nějaké latentní třídy – $P(C|c)$, kde C je latentní třída, do níž jednatel skutečně náleží, a c je třída, do níž byl jednatel modelem zařazen.

Pro interpretaci jsou někdy vhodnější obráceně podmíněné pravděpodobnosti $P(c|C)$, tedy pravděpodobnost, s jakou je jedinec modelem klasifikován do třídy k , je-li ve skutečnosti členem třídy k . Bayesovým vzorcem lze $P(C|c)$ snadno převést na $P(c|C)$, protože marginální pravděpodobnosti $P(C)$ jsou odhadovány modelem a apriorní $P(c)$ jsou relativní četnosti těch, kdo mají nejvyšší P_{ik} pro tu kterou třídu.

$$P(c = k|C = K) = \frac{P(C = K|c = k)P(c = k)}{P(C = K)}$$

$P(c|C)$ se někdy nazývají **klasifikační pravděpodobnosti** a Mplus je v případě modelů s latentními třídami standardně vypisuje do výstupu pod nadpisem *Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)*. Příklad v tabulce II.4 je ze stejného modelu jako předchozí tabulka, takže pokud náleží jedinec do 1. latentní třídy, je pravděpodobnost, že jej tam daný model zařadí (= bude mít nejvyšší posteriorní pravděpodobnost právě v této třídě) rovna 0,98. To je poměrně vysoká pravděpodobnost, která je dána tím, že četnost 1. latentní třídy je v tomto modelu vysoká ($P(C=1) = 0,85$) stejně jako apriorní pravděpodobnost $P(c=1) = 0,896$. Pro malou 2. třídu ($P(C=2) = 0,15$) již je správná klasifikace mnohem méně pravděpodobná $P(c=2|C=2) = 0,62$, protože 1. třída má vysokou apriorní pravděpodobnost.

Tabulka II.4. Klasifikační pravděpodobnosti ve výstupu Mplus.

Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)		
	1	2
1	0.982	0.018
2	0.376	0.624

Z průměrných posteriorních pravděpodobností pro jednotlivé latentní třídy $P(C|c)$ je tedy vidět, jak dobře funguje klasifikace latentních tříd bez přihlídnutí k jejich mohutnosti. Klasifikační pravděpodobnosti ještě navíc vezmou v potaz mohutnost latentních tříd. V obou případech považujeme klasifikaci za dobrou, když jsou v tabulkách těchto pravděpodobností v diagonále vysoké hodnoty (0,8 a vyšší) a mimo diagonálu nízké hodnoty. Podle tohoto kritéria mohou být v rámci jednoho modelu s více latentními třídami některé třídy „dobře oddělené“ a jiné ne.

Jasnost klasifikace vyjádřená posteriorními pravděpodobnostmi a klasifikačními pravděpodobnostmi lze ještě více shrnout do jednoho čísla, statistiky zvané **entropie**. Ta je v souladu s obsahem této veličiny napříč přírodními vědami vyjadřuje míru uspořádanosti,

ale zatímco v přírodních vědách znamená vysoká entropie nízkou míru uspořádanosti systému, v kontextu klasifikace je entropie počítána tak, že její vysoké hodnoty znamenají vysokou míru uspořádanosti, jasnosti klasifikace, a nízké hodnoty nejasnost klasifikace, kdy jednotlivec může se podobnou pravděpodobností náležet do více (všech) latentních tříd. Mplus počítá hodnotu entropie podle následujícího vzorce (Asparouhov & Muthén, 2014b)

$$E = 1 + \frac{1}{N \ln(k)} \sum_{i=1}^N \sum_{k=1}^K P(c = k|U_i) \ln(P(c = k|U_i)).$$

kde c je třída, do níž je jednotlivec modelem klasifikován, K je počet tříd, N velikost vzorku a U_i je vektor indikátorů c (např. I, S v případě růstových modelů). Jde tedy o sumu všech posteriorních pravděpodobností všech jednotlivců vynásobených jejich přirozeným logaritmem, která je normalizovaná přibližně na rozpětí 0-1.

Mplus umožňuje spočítat i entropii, kterou lze připsat na vrub jednotlivým manifestním indikátorům (*univariate entropy*) uvedením příkazu ENTROPY; do sekce OUTPUT (Asparouhov & Muthén, 2014b). Jde o míru, v níž se latentní třídy liší v jednotlivých měřeních, na nichž se zakládá růstová křivka. V příkladu výstupu Mplus z modelu, který používám v této sekci, je vidět, že k vysoké entropii přispívá nejvíce měření z 15 let. V tomto věku se od sebe růstové křivky obou latentních tříd v modelu od sebe nejvíce lišily (více viz kapitola III.).

Tabulka II.5. Univariační entropie ve výstupu Mplus.

CLASSIFICATION QUALITY				
Entropy	0.763			
Univariate Entropy				
S2LN_13	S2LN_15	S2LN_17	S2LN_19	
0.472	0.719	0.533	0.448	

Entropie nese velmi odlišnou informaci od ukazatelů založených na shodě modelu s daty. I model s velmi nízkou entropií může mít velmi dobrou shodu s daty; zvláště u modelů směsí (GMM), v nichž mají růstové křivky nenulovou variabilitu uvnitř latentních tříd. Třídy se kvůli tomu „překrývají“ a případy v těchto přechodových oblastech parametrů je obtížné klasifikovat.

S narůstajícím počtem tříd obvykle (ne však nutně) dochází k poklesu entropie. Při volbě optimálního modelu je tedy potřeba hledat kompromis mezi počtem a s ním spojenou specifičností latentních tříd a dostatečnou entropií. Pro žádoucí hodnoty entropie nemáme pevná kritéria. Dolní limit (0) i horní limit (1) jsou v praxi neobvyklé. Obvykle se považují za dostatečné hodnoty entropie nad 0,7 (Muthén, 2004).

Obecněji o separaci tříd píše (Depaoli, 2013) a podobně jako (Tein et al., 2013) používá jako měřítko teoretické oddělenosti latentních tříd Mahalanobisovu vzdálenost průměrných růstových parametrů mezi třídami.

SHRNUTÍ

Nylund et al. (2007) publikovali velkou Monte Carlo studii, kde porovnávali běžný LRT s BLRT, LMR (Lo, Mendell, & Rubin, 2001) a informačními kritérii. Vedle LCA, LPA a GMM modelů to aplikovali i na relativně řídko používaný Factor Mixture Analysis (FMA) model (Lubke & Muthén, 2005). Téměř ve všech scénářích fungoval nejlépe BLRT a BIC, a to jak z perspektivy chyby 1. typu, tak z perspektivy síly testu. Solidně funguje i CAIC a LMR. LMR lze využít v počátcích analýzy pro jeho menší výpočetní náročnost oproti BLRT. Vzorky kolem 200 dávají podstatně menší jistotu než 500. Mezi 500 a 1000 už není tak velký skok. Uvedené

odhady potřebné velikosti vzorku pro dostatečně silný BLRT test jsou však pouze vedlejšími produkty simulací na určitých prototypických modelech. O power analýzu BLRT testu se na analýze latentních tříd pokusili (Dziak, Lanza, & Tan, 2014). Navrhli pro tento účel několik metrik velikosti účinku, tedy odlišnosti, separovanosti tříd, z nichž dvě měly jasný vztah s velikostí účinku – modifikaci Cohenova w pro chí-kvadrát testy nad kontingenčními tabulkami a koeficient velikosti účinku založený na Kullbackově-Leiblerově divergenci. I pro LCA model, který je podstatně jednodušší než GMM modely, je pro 80% sílu BLRT testu potřeba řádově tisíce účastníků při malé velikosti účinku ($w = 0,1$) a stovky při střední velikosti účinku ($w = 0,3$). Power analýzu BLRT testu prezentují také Tekle, Gudicha, & Vermunt (2014) s podobnými závěry – pro dostatečnou sílu testu je potřeba v jednodušších modelech 200 ve složitějších až 2000 případů. I oni ale své simulace zakládají na jednoduchém LCA modelu s dichotomickými indikátory. Pro GMM, vzhledem jejich složitosti je dobré výzkumy plánovat spíše na horní hranici těchto doporučení.

6. Zařazení kovariátů do růstového modelu

Identifikací, respektive enumerací tříd a jejich popisem a interpretací analýza samozřejmě nekončí. Charakteristika, jejíž vývoj modelujeme, se vyvíjí v kontextu dalších intra-individuálních i environmentálních proměnných. Tyto proměnné – kovariáty - mohou hrát v našich úvahách různé role, které v kontextu lineárních modelech nazýváme tradičními termíny mediace, moderace, podmínené efekty a podobně (McLarnon & O'Neill, 2018).

1. Můžeme chtít postupovat explanačně či deskriptivně a hledat souvislosti mezi členstvím v latentní třídě a hodnotami kovariátů. Pokud latentní třídy odpovídají reálně existujícím subpopulacím, lze předpokládat, že členství v latentní růstové třídě může mít s kovariáty vztahy, které se liší od korelací průřezových, jednorázových měření a které nám umožní lépe porozumět zkoumanému jevu, odhalit dosud skryté vztahy.
2. Můžeme uvažovat „psychometricky“ a „obávat se“, že kovariáty mohou narušovat předpoklady růstového mixture modelu, zasahovat do procesu odhadu počtu tříd a parametrů jednotlivých tříd (Nylund-Gibson & Masyn, 2016, Li & Hser, 2011). To by pak mohlo omezit „reálnost“ identifikovaných subpopulací nebo zkreslenost jejich parametrů v důsledku nezohlednění nebo špatně specifikovaného zohlednění kovariátu.

uvádějí čtyři různé způsoby, jimiž může být kovariát specifikován jako součást modelu s latentními třídami:

1. Nepřímý efekt. Kovariát má efekt na latentní třídu (tedy jeho hodnoty činí některé třídy pravděpodobnější než jiné, obvykle jako v multinomiální regresi) a jejím prostřednictvím má nepřímý efekt na indikátory latentní třídy (průsečík, směrnice apod).
2. Kovariát může mít přímý efekt na jeden nebo více indikátorů latentní třídy.
3. Kovariát může mít jak přímý, tak nepřímý efekt
4. Kovariát může mít přímý a nepřímý efekt, přičemž přímý efekt může být moderován latentní třídou

I když je myslitelný a v literatuře někdy používaný jednoduchý postup, který známe ze shlukových analýz, kdy si do dat uložíme nejpravděpodobnější třídu každého jednotlivce a tuto proměnnou pak používáme jako faktor či multinomiální závislou v manifestních

lineárních modelech, není takový postup zdaleka optimální. Hlavním důvodem je to, že při práci s "nejpravděpodobnější třídou" se ztrácí informace o tom, s jakou jistotou je ten který jedinec kategorizován. Pravděpodobnost pro jedince nejpravděpodobnější třídy totiž může být nepříjemně nízká (třeba 0,4). Tento problém lze sice částečně řešit v následných analýzách odfiltrováním těch, kdo mají nejvyšší pravděpodobnost třídy menší než nějakou zvolenou hladinu, např. 0,7, ale to je jen provizorní řešení, které při relativně nízké hodnotě entropie a větším počtu tříd (>3) může znamenat nepříjemnou redukci vzorku.

Lanza, Tan, & Bray (2013) také uvádějí, že efekt latentní třídy na závislé proměnné může být těmito *classify-analyze* postupy podhodnocený. Dalším důvodem je to, že tímto způsobem nelze řešit otázky spojené s rolí kovariátu v samotné enumeraci tříd. Vhodnější se tedy jeví využít některý z „model-based“ přístupů, kdy jsou kovariáty nějakým způsobem začleněny do modelu latentních tříd.

JEDNOKROKOVÝ POSTUP

Prediktor/kovariát lze do GMM modelu zařadit přímo v roli prediktoru latentních tříd či manifestních proměnných a odhadovat jeho efekt společně s ostatními parametry GMM modelu. To má vést k přesnějšmu odhadu velikosti účinku kovariátu.

Muthén (2004) uvádí tři role, v nichž lze kovariát začlenit do GMM modelu: 1) v roli prediktoru latentních tříd, což má podobu multinomiální regrese (*between-class effect*), 2) v roli prediktoru růstových parametrů (*within-class effect, mixture regression*) a (3) v roli distálních projevů (outcomes) latentních tříd. Může se vyskytnout i ve více rolích zároveň (Nylund-Gibson & Masyn, 2016).

V rámci obecného SEM modelu může být kovariát samozřejmě nejen manifestní, ale i latentní. U *within-class efektu* přibývá k dříve komentovaným volbám ohledně toho, co se mezi třídami smí a nesmí lišit, další volba, a to je rozhodnutí, zda odhadovat efekt v každé třídě zvlášť, nebo zda jej odhadnout jako stejný napříč skupinami. Přirozenější je asi nechat jej jiný, když nám mají kovariáty pomoci porozumět smyslu tříd. Within-class efekty samozřejmě nejsou možné u LCGA modelů, kde nemají růstové parametry žádný rozptyl.

Modelování efektů členství ve třídě záleží na tom, jestli je závislá proměnná dichotomie nebo spojitá proměnná. Pro dichotomii Mplus počítá logistickou regresi - odhaduje prahy pro každou třídu, přičemž prahy udávají proporci distálního projevu v dané třídě (menší než 0, převažují projevy kódované 1, větší než 0, převažují projevy kódované 0). Pro spojitou distální je to totéž, pouze se místo prahů počítají přímo průměry pro každou třídu. V zásadě se tedy odhadují prahy nebo průměry distální proměnné pro každou třídu a model test je srovná (Asparouhov & Muthén, 2015).

Bakk, Oberski, & Vermunt (2016) uvádí, že pokud rozložení závislé proměnné není uvnitř jednotlivých latentních tříd normální, může zařazení kovariátu do modelu změnit parametry latentních tříd, ale i jejich počet. Jak bylo uvedeno výše, může to být z hlediska enumerace latentních tříd žádoucí. Často bychom však rádi znali vztah mezi členstvím v latentní třídě a kovariátem, aniž by kovariát hrál roli v samotné definici latentních tříd. Pak jde o nepříjemný problém.

Tabulka II.7. Srovnání parametrů dvou latentních tříd užívání návykových látek v modelech lišících se zařazením kovariátu pohlaví.

	Nepodmíněné řešení		Pohlaví jako prediktor třídy		Pohlaví jako prediktor růstových parametrů	
n1	497		495		492	
n2	58		60		63	
p1	0,895		0,892		0,886	
p2	0,105		0,108		0,113	
Entropie	0,763		0,764		0,766	
c1		se		se		se
S2LN_15	1,769	0,155	1,77	0,155	1,755	0,156
S2LN_17	4,862	0,161	4,861	0,162	4,843	0,161
c2						
S2LN_15	5,515	0,557	5,52	0,565	5,503	0,547
S2LN_17	6,412	0,584	6,417	0,588	6,409	0,569
c1						
I	3,499	0,125	3,499	0,125	3,333	0,182
S	0,956	0,037	0,956	0,037	1,001	0,053
c2						
I	5,713	0,479	5,716	0,476	6,2	0,613
S	1,032	0,121	1,031	0,122	0,973	0,14

Poznámka. n - četnost respondentů klasifikovaných do třídy, p - relativní četnost třídy v populaci, S2LN_15 a _17 - odhadované časové báze, I a S - průsečík a směrnice.

Například u modelu růstu užívání návykových látek v adolescenci (kouření a konzumace alkoholu) ze studie ELSPAC (viz kapitola III.) je mezi nepodmíněným GMM modelem se třemi třídami (S2_GM2_3) a modelem do něhož bylo zařazeno pohlaví v roli prediktoru členství v třídě řada rozdílů v parametrech odhadnutých tříd a v entropii. Přitom v modelu se dvěma třídami, jehož odhad je robustnější, entropie vyšší a BLRT rozdílu mezi 2- a 3-třídním modelem je nesignifikantní, jsou rozdíly v parametrech tříd mezi nepodmíněným modelem a modelem s kovariáty zanedbatelně malé. To ale neznamená, že tam nejsou. V modelu s pohlavím v roli prediktoru členství ve třídě se lišila velikost tříd o 2 respondenty, v modelech s pohlavím v roli prediktoru růstových parametrů o 5 respondentů. Odhadnuté růstové parametry jednotlivých tříd se mezi modely s a bez kovariátu lišily o méně než setinu směrodatné chyby parametru. To není vzhledem k $N = 555$ mnoho, ale dokladuje to drobnou nestabilitu mixture modelů, kterou je potřeba mít na paměti. V případě 2- i 3-třídního modelu nemá pohlaví na členství ve třídě vliv. Pokud by však mělo, nabízela by se otázka, zda se pohlaví přímo nepodílí na vzniku adolescentních tříd užívání návykových látek (namísto predikování) a zda jej tedy záměrně nezařadit již na začátku procesu hledání mixture růstového modelu.

TŘÍKROKOVÝ PŘÍSTUP

Kvůli tomu, že zařazení prediktorů může významně ovlivnit parametry latentních tříd, navrhuje Vermunt (2010) oddělení odhadu parametrů latentních tříd od modelování vztahu pravděpodobnostního členství v latentní třídě s kovariáty, ať již v roli prediktorů, či závislých. Tento postup se skládá z následujících tří kroků (a proto se v literatuře označuje jako tříkrokový přístup – *3-step approach*):

1. Odhadneme nepodmíněný mixture model a necháme software vyexportovat do datového souboru nejpravděpodobnější třídu pro každého jednotlivce. V Mplus k tomu slouží sekce „*SAVEDATA:*“ v níž specifikujeme jméno souboru, do něhož se má nejpravděpodobnější třída (a další proměnné) zapsat (*FILE=*) a vyžádáme se uložení požadovaných údajů („*SAVE=CPROB;*“).

2. Vytvoříme "repliku" kostry nepodmíněného mixture modelu z nejpravděpodobnější třídy, což je vlastně manifestní podoba původně latentní třídy a z informace o klasifikační nejistotě. V modelu je tak pevně specifikováno, kdo a s jakou pravděpodobností spadá do které latentní třídy. V tomto modelu již nejsou zahrnuty indikátory latentních tříd (tedy v případě GMM parametry růstových přímek) – zůstává pouze pravděpodobnostní členství ve třídě. Pro informace o charakteristikách té které třídy se musíme podívat na parametry modelu odhadnutého v prvním kroku.

Pro vytvoření této „repliky“ v Mplus pracujeme s datovým souborem, který obsahuje nejméně jednu proměnnou – tu, která udává nejpravděpodobnější členství ve třídě. Obvykle však již v této fázi chceme mít v datovém souboru zahrnuty i všechny kovariáty, jejichž efekt bychom chtěli modelovat. Proměnná udávající nejpravděpodobnější třídu (např. T) je v Mplus specifikována jako nominální proměnná („*NOMINAL = T;*“). Informace o nejistotě ohledně členství v latentní třídě (obvykle C) se do modelu zahrne nikoli z dat, ale je přímo specifikována ve specifikaci jednotlivých latentních tříd. Požadovaného efektu je dosaženo tak, že se v jednotlivých třídách pevně fixuje pravděpodobnost, s jakou nabývá T (nejpravděpodobnější třída) různých hodnot. V modelu, kde je klasifikační nejistota minimální bychom očekávali, že v první latentní třídě (C#1) bude proměnná T udávající vyexportovanou nejpravděpodobnější třídu nabývat s vysokou (téměř 100%) pravděpodobností své první hodnoty (T#1 je v jazyce Mplus první(nejnižší) hodnota nominální proměnné T) a s velmi nízkou pravděpodobností (téměř 0 %) hodnot ostatních. Čím více je přítomno klasifikační nejistoty, tím vzdálenější jsou vysoké i nízké pravděpodobnosti od svého ideálu. V jazyce Mplus se tyto pravděpodobnosti zadávají ne přímo, ale v podobě logitu členství v dané třídě oproti referenční třídě (tedy logaritmus poměru pravděpodobnosti členství v dané třídě ku pravděpodobnosti v referenční třídě). Mplus používá jako referenční třídu poslední třídu a požadované logity poskytuje ve výstupu v tabulce pojmenované *Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)*. Logity v tabulce jsou spočítány z klasifikačních pravděpodobností členství ve třídě uvedených v tabulce, která jí ve výstupu Mplus bezprostředně předchází „*Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)*“. Ve výsledku tak do syntaxu uvádíme například pro model se třemi latentními třídami:

```
%C#1%  
[T#1@6]; [T#2@0.4];  
%C#2%  
[T#1@0.4]; [T#2@6];  
%C#3%  
[T#1@-4]; [T#2@-4];
```


Znamená to, že ve třídě jedna je pravděpodobnost $T=1$ e^6 krát vyšší (403x vyšší) než pravděpodobnost $T=3$ a $T=2$ je $e^{0.4}$ krát vyšší (1,5x vyšší) než $T=3$. Z toho plyne že $P(T=1 | C=1)$ je velmi vysoká a $P(T=2 | C=1)$ a $P(T=3 | C=1)$ jsou podobně velmi nízké, což znamená velmi nízkou klasifikační nejistotu.

Na konci 2. kroku tedy máme „repliku“ latentních tříd a můžeme si zkontrolovat, že všechny klasifikační pravděpodobnosti jsou v replice stejné, jako v původním nepodmíněném mixture modelu. A díky fixování logitů se pravděpodobnostní členství v latentní třídě obvykle nezmění, když do modelu přidáme další proměnné.

3. Do repliky modelu přidáme kovariáty, které budou modelovány stejně jako v jednokrokovém postupu. Prediktory členství ve třídě budou pomocí multinomické regrese predikovat rozdíl logitu členství ve třídě spojený s jednotkovým rozdílem hodnoty prediktoru. Podobně je efekt členství ve třídě modelován pomocí odhadu průměru závislé proměnné v jednotlivých třídách. Vzhledem k tomu, že v modelu již nejsou indikátory členství ve třídě, není tímto postupem možné modelovat i within-class efekty prediktorů, tedy efekty na hodnoty parametrů růstových křivek uvnitř jedné nebo více tříd.

Tento tříkrokový postup lze realizovat ručně, ale některé statistické programy nabízí automatizaci tohoto postupu. V kontextu Mplus jsou kovariáty zařazené ve třetím kroku nazývány *auxiliary variables* a nabízí řadu možností, jak je modelovat (Asparouhov & Muthén, 2014a) (Asparouhov & Muthén, 2015). Mplus nabízí příkaz AUXILIARY, kterým se v sekci VARIABLE: mohou definovat kovariáty, pro které se má realizovat tříkrokový postup. Tímto způsobem lze specifikovat jako prediktory, tak outcomes. Bohužel je nyní možné modelovat buď prediktory členství v latentní třídě, nebo efekty členství v latentní třídě na nějaké outcomes, ale ne obojí najednou. Role kovariátu je v příkazu AUXILIARY specifikována v závorkách. (R3STEP) tak přiřazuje kovariátu roli prediktoru členství v latentní třídě. Například AUXILIARY=(R3STEP) VAR1; instruuje Mplus, aby realizoval tříkrokový postup s proměnnou VAR1 v roli prediktoru členství v latentní třídě. Pro přiřazení role závislé proměnné nabízí Mplus dvě možnosti: (DU3STEP) a (DE3STEP). Volba (DE3STEP) způsobí, že rozložení závislé proměnné napříč latentními třídami bude modelováno tak, že rozptyly závislé budou modelovány jako invariantní napříč třídami (dependent-equal). Budou se tedy lišit pouze průměry a test jejich rozdílů bude tedy ekvivalentní t-testu s předpokladem homoskedacity. Volba (DU3STEP) umožní rozptylům závislé proměnné, aby se lišily napříč latentními třídami. Asparouhov & Muthén (2014a) uvádí, že (DE3STEP) je určena pouze pro situace s malými třídami, kdy mohou nastat estimační obtíže v důsledku rozptylů pohybujících se blízko 0. Výchozí volbou by tedy měla být (DU3STEP).

Diallo & Lu (2017) na aktuální rozsáhlé simulaci srovnávající jednokrokový a tříkrokový přístup uvádí, že obecně jim v simulacích vycházel jako vhodnější výchozí volba jednokrokový přístup, protože úspěšnost tříkrokového přístupu závisela na specifikaci mixture modelu v prvním kroku, na velikosti účinku kovariátů, typu kovariátů a dalších okolnostech.

Jak píše Asparouhov & Muthén (2014a) tříkrokový přístup nezaručuje, že nemůže dojít ke změně parametrů latentních tříd (*class shift*). K posunu tříd bohužel někdy dochází. Výhodou použití DU3STEP a DE3STEP je, že Mplus pak automaticky kontroluje, jestli nedošlo v důsledku zapojení kovariátů ke změně pravděpodobnostních členství v latentních třídách. V případě, že k tomu dojde Mplus o tom ve výsledcích informuje a efekty latentní třídy na závislou v takovém případě nevypíše.

BCH PŘÍSTUP

Vedle standardního tříkrokového postupu popsaného výše navrhl Vermunt (2010) ještě jeden tříkrokový postup vycházející z metodiky, kterou publikovali Bolck, Croon, & Hagenaars (2004). Podle jmen autorů je postup pojmenovaný BCH, byť v kontextu GMM (a Mplus) se touto zkratkou obvykle míní Vermuntova modifikace tohoto postupu Vermunt (2010). Stejně jako u standardního tříkrokového postupu začíná analýza odhadem modelu měření latentních tříd a následnou pravděpodobnostní klasifikací do tříd. Následně specifikujeme multigroup SEM model, kde figuruje třída jako proměnná definující skupiny. Klasifikační nejistota je zohledněna vážením jednotlivých případů, kdy (velmi zjednodušeně) mají případy s vysokou pravděpodobností členství ve třídě vyšší váhu než případy s nízkou pravděpodobností členství ve skupině. Tento systém vah je poměrně složitý a váhy mohou v případě nízké entropie (vysoké klasifikační nejistoty) nabývat i záporných hodnot. Proto se obvykle nesetkáváme s manuálním provedením tohoto postupu a spoléháme na software, v němž je postup naprogramovaný (Mplus, Latent Gold).

Výhody BCH postupu mají být trojí. Jednak má vést k lepším odhadům efektu třídy z hlediska zkreslení i MSE, a to i když se rozptýl závislé proměnné napříč třídami výrazně liší (Asparouhov & Muthén, 2015). Druhou výhodou má být to, že tím, že v druhém kroku již se neodhaduje mixture, resp. latent class model, již nemůže dojít ke změně parametrů tříd. Třetí, a z mé perspektivy nejzásadnější výhodou je to, že multigroup SEM model, jímž odhadujeme efekt latentní třídy na závislé proměnné, nemusí být jen tím triviálně jednoduchým modelem modelujícím průměry ve skupinách odpovídajících latentním třídám. Může jít v zásadě o jakýkoli SEM model s latentními proměnnými, pro který si je možné položit otázku, zda se liší parametry modelu v závislosti na latentních třídách. To otevírá velké množství zajímavých hypotéz.

Nevýhodou BCH přístupu je podle Asparouhova a Muthéna (2015) to, že negativní váhy v případě nízké entropie mohou někdy vést i k negativním odhadům rozptylů, tedy neplatným řešením. Protože jsou podle Bakka & Vermunta (2016) odhady průměrů nezávislé na odhadech rozptylů, lze takovou situaci řešit omezením rozptylů na stejnou hodnotu napříč latentními třídami.

Mplus nabízí dvě implementace BCH postupu. Ten, který pouze odhaduje rozdíly průměrů napříč skupinami odpovídající latentním třídám, nazývá Asparouhov *automatický BCH*. Automatický je v tom smyslu, že uživatel definuje model latentních tříd či mixture model a v syntaxu uvede, pro jaké kovariáty se mají spočítat průměry v latentních třídách (AUXILIARY = COV1(BCH);). Mplus pak realizuje jednotlivé kroky automaticky a prezentuje je v jednom výstupu. Tento postup předpokládá, že závislá proměnná je spojitá; pro kategorické závislé je nutné využít buď manuální BCH, nebo LTB přístup (viz níže). Manuální BCH umožňuje modelovat efekt latentních tříd na jakýkoliv námi specifikovaný SEM model, což vyžaduje spuštění dvou modelů. Jedním se odhadne model měření latentních tříd či směsí a vyexportuje se datový soubor obsahující členství ve třídách a váhy. V druhém se specifikuje multigroup SEM model, na který mají latentní třídy hypotetizovaný efekt, s tím, že váhy jednotlivých případů zohlední klasifikační nejistotu.

Příklad podle Asparouhov & Muthén (2015)

```
Model 1:
Variable:
Names=U1-U10 Y X;
Categorical = U1-U10;
Classes = C(3);
Usevar=U1-U10;
Auxiliary=Y X;
Data: file=manBCH.dat;
Analysis: Type = Mixture;
Savedata: File= manBCH2.dat;
Save=bchweights;
```

```
Model 2:
Variable: Names = U1-U10 Y X W1-W3 MLC;
Usevar are Y X W1-W3;
Classes = C(3);
Training=W1-W3 (bch);
Data: file=manBCH2.dat;
Analysis: Type = Mixture; Starts=0;
Estimator=mlr;
Model:
%overall% Y on X;
%C#1% Y on X;
%C#2% Y on X;
%C#3% Y on X;
```

LTB PŘÍSTUP

Lanza et al., (2013) navrhli postup, který lze použít pouze pro modelování efektu členství v latentní třídě na závislé proměnné (outcomes), tak abychom zařazením kovariátu neovlivnili individuální pravděpodobnostní členství v latentních třídách či celý model měření latentních tříd (LBT, Lanza-Tan-Brey, nebo jen „Lanza“ přístup). Vyšli z toho, že zařazení prediktorů nemá na modely takový efekt, jako zařazení proměnných závislých na latentní třídě. Jejich postup je tedy v principu takový, že závislou proměnnou do modelu zařadí jako prediktor a z odhadnutých parametrů modelu pomocí Bayesova vzorce zpětně dopočítají rozložení (průměry) závislé proměnné v jednotlivých třídách. Předpokladem tohoto přístupu je, že indikátory tříd jsou podmíněně nezávislé na kovariátu, zohledníme-li efekt tříd. Tento postup je překvapivě účinný, pokud platí předpoklad log-linearity vztahu mezi kovariátem a závislou. Tento předpoklad je však často porušen, například prostřednictvím heteroskedacity rozptylů kovariátu napříč latentními třídami a potom jsou negativně ovlivněny nejen odhady směrodatných chyb, ale i odhady samotných průměrů jsou zkreslené, a to tím více, čím se rozptyly napříč třídami liší (Bakk & Vermunt, 2016). Tento a další limity se snaží řešit Bakk et al. (2016) pomocí zařazení kvadratického členu do predikce třídy kovariátem a rozdělením postupu do tří kroků. Podle MC simulací provedených na jednoduchých LCA a LPA modelech produkoval LBT přístup poměrně malou míru zkreslení ve srovnání prostým kategorizováním do nejpravděpodobnější třídy (Collier & Leite, 2017).

Mplus umožňuje použít LTB postup pomocí specifikace auxiliary proměnných jako (DCON) v případě spojitých a (DCAT) v případě kategorických závislých proměnných. (Asparouhov & Muthén, 2014a) uvádí, že se v některých detailech odhadu odlišují od Lanza, Tan & Brey (2013), zejména ve způsobu odhadu směrodatných chyb (Mplus je bootstrapuje), ale výsledky odhadu jsou srovnatelné.

Shrnutí

Rozhodování ohledně způsobu modelování vztahu latentních tříd (komponent směsí) s různými kovariáty ve své složitosti trochu připomíná volbu post-hoc testu u analýzy rozptylu. Asparouhov & Muthén (2015) uvádí užitečný přehled přístupů dostupných v Mplus. Zatímco jednokrokové postupy umožňují velkou flexibilitu s kovariáty v roli prediktoru, závislé i vnitro-třídních kovariátů, jsou zároveň ohroženy všemi nebezpečími, s nimiž se můžeme v této oblasti setkat – se změnou parametrů či počtu tříd, zkreslením odhadů parametrů i odhadů jejich směrodatných chyb. Pokud si zkontrolujeme, zda nedošlo k posunu tříd, je jeho vhodnou náhradou (či kontrolou) manuální třídkrokový přístup.

Případné vnitrotřídní kovariáty je však nutné mít již v modelu v prvním kroku. Pokud se potřebujeme zaměřit pouze na prediktory či pouze na efekt tříd na závislé proměnné, je vhodné využít automatické optimalizované přístupy – R3STEP pro prediktory, BCH pro spojité závislé a DCAT pro kategorické závislé.

Tabulka II.8. Možnosti modelování vztahu latentních tříd s kovariáty (podle Asparouhov & Muthén, 2015)

Přístup	Klíčové slovo v příkazu AUXILIARY (...)	Typ kovariátu	Role kovariátu	Poznámky
Jednokrokový		spojitý i kategorický	závislá i prediktor	Flexibilní, ale s nebezpečím posunu tříd a zkreslení odhadů
Tříkrokový manuální dle Vermunta (2010)		spojitý i kategorický	závislá i prediktor	Možnost posunu tříd.
Pseudo-class (PC) podle Wang et al. (2005)	(E)	spojitý	závislá	Zkreslené výsledky, lepší je DU3STEP
Pseudo-class (PC) podle Wang et al. (2005)	(R)	spojitý i kategorický	prediktor	Zkreslené výsledky, lepší je R3STEP
Tříkrokový přístup dle Vermunta (2010)	(R3STEP)	spojitý	prediktor	Doporučená metoda pro prediktory
Tříkrokový přístup dle Vermunta (2010) a Asparouhova a Muthéna (2014)	(DU3STEP)	spojitý	závislá	Pokud Mplus nevypíše výsledky, došlo k posunu tříd a je na místě použít BCH
Tříkrokový přístup dle Vermunta (2010) a Asparouhova a Muthéna (2015)	(DE3STEP)	spojitý	závislá	Obměna DU3STEP - rozptyly v třídách odhadovány jako stejné. Jen pro nouzové použití v případech potíží s konvergencí.
Automatický BCH přístup podle Bakk a Vermunt (2014)	(BCH)	Spojité	závislá	Preferovaná metoda pro spojité závislé. Při nízké entropii může dojít k podhodnocení SE. Mplus nabízí podporu i pro manuální verzi umožňující jakýkoli závislý model.
LTC přístup podle Lanza et al. (2013) a Asparouhova a Muthéna (2015)	(DCON)	spojitý	závislá	Problémy při nízké entropii (<0,6) a nestejných rozptylech napříč třídami. Vhodnější je BCH a DU3STEP
LTC přístup podle Lanza et al. (2013)	(DCAT)	kategorický	závislá	Preferovaná metoda pro kategorické závislé. Nedochází k posunu tříd.

Výše popsany postup hledání latentních tříd či komponent směsi růstových křivek je poměrně náročný. I když jsem jej popsal sekvenčně, je zřejmé, že reálně je potřeba se pohybovat mezi jednotlivými kroky tam a zpátky. V každém kroku je k dispozici řada voleb, které mohou mít teoretický podklad a praktické konsekvence v dalším kroku. Nepodmíněný růstový LGC model, který se zdá velmi vhodným popisem dat, se může ve fázi hledání latentních tříd ukázat parametricky příliš náročným, popřípadě příliš složitým. Tak jako se

rozložení jednotlivých růstových parametrů skládá v mixture modelu z několika třídně-specifických rozložení, se i průměrná nelineární růstová trajektorie pro celou populaci může skládat z lineárních třídně-specifických trendů. Může se také stát, že první modely s latentními třídami poukážou na charakteristiky individuálních růstových křivek, kterých si analytik zatím nevšiml a má potřebu je zohlednit v LGC modelu. Podobně jako u faktorové analýzy může celá procedura nakonec vyústit v to, že data představu diskrétních subpopulací vůbec nepodporují. Při tom je potřeba zvažovat jak statistická kritéria vhodnosti či přímo přípustnosti odhadnutých parametrů jednotlivých modelů, ale samozřejmě i jejich teoretickou smysluplnost či heuristický nebo praktický užitek.

Pro představu tohoto procesu předkládám v následujících dvou kapitolách dvě analýzy latentních tříd růstových křivek. Jsou popsány v detailu, který není v běžných empirických studiích běžný a vlastně ani příliš užitečný. Namísto prezentování finálního modelu či několika málo modelů, prezentuji téměř všechny zvažované a odhadované modely, aby byly vidět větší či menší dopady všech možných voleb popsaných v předchozím textu. První analýza se věnuje vývoji jednoho aspektu rizikového chování – užívání běžných návykových látek – v adolescenci mezi 13. a 19. rokem. Tato analýza je nejjednodušším příkladem – vychází z pouhých čtyř měření v pravidelných rozestupech, pracuje s manifestními skóry a jde o lineární model. Druhá analýza se věnuje vývoji prožívané autonomie v mladém dospělosti mezi 18. a 28. rokem a představuje komplexnější podobu analýzy latentních tříd růstu. Pracuje s mnoha měřeními prožívané autonomie, které se odehrály v různě dlouhých intervalech během čtyř let, přičemž do výzkumu vstupovali lidé ve věku 18–24 let, takže individuální růstové křivky začínají a končí v různých věcích. V datech pro celé věkové rozpětí 18–28 je tak mnoho designem studie daných chybějících hodnot. Vzhledem k velikosti vzorku a kvalitě použité měřicí škály prožívané autonomie bylo možné aplikovat i růstový mixture model druhého řádu, tedy modelovat vývoj latentních skóre autonomie.

III. ANALÝZA HETEROGENITY VÝVOJOVÝCH KŘIVEK RIZIKOVÉHO CHOVÁNÍ V ADOLESCENCI

V rámci longitudinální studie ELSPAC (Ježek, Macek, Lacinová, 2011) jsme mimo jiné sledovali i vývoj různých chování, která mohou představovat nějakou míru rizika pro úspěšný psychosociální vývoj v adolescenci. K získání dat byl použit self-reportový dotazník rizikového chování (Širůček, Širůčková, 2008) zachycující potenciálně riziková chování v pěti oblastech – školní problémy, užívání návykových látek, delikvence, agrese a konflikty v rodině. Dotazník byl administrován čtyřikrát ve 13, 15, 17 a 19 letech. Protože administrace proběhla v každém věku jednorázově a věkové rozpětí vzorku bylo 18 měsíců, skutečný věk jednotlivých účastníků se v jednotlivých vlnách mohl od středního věku odlišovat až o 9 měsíců. Přesto jsme data analyzovali tak, že čas reprezentovala vlna měření, nikoli přesný věk účastníka výzkumu. Beranová, Ježek, Širůček (2011) prezentují modely růstových křivek popisující lineární růst v pěti oblastech rizikového chování a souvislost růstu s pohlavím, vztahem s rodiči, vztahem s vrstevníky a styly identity. Heterogenita individuálních růstových křivek však byla vysoká, a tak GCM modely představovaly velmi strohý model vývoje rizikového chování. V této kapitole bych chtěl navázat na analýzy prezentované v Beranová, Ježek, Širůček (2011) a prozkoumat, zda lze popsat vývoj rizikového chování v adolescenci jednou variabilní vývojovou křivkou, nebo zda je vhodnější uvažovat o tom, že různé skupiny (typy) adolescentů se vyvíjí různě.

V analýze jsem se zaměřil na škálu **užívání návykových látek**, která je součtovou škálou kombinující self-reportované užívání piva nebo vína, tvrdého alkoholu, cigaret, a marihuany. Všechny čtyři položky měly šestibodovou odpověďovou škálu od nikdy (1) po několikrát denně (6). Součtová škála má uspokojivou vnitřní konzistenci (Cronbachovy alfy jsou od 13 do 19 let 0,62, 0,77, 0,73, a 0,73) a její hodnoty byly vypočítány jako průměr odpovědí. Protože rozložení této proměnné bylo ve všech věcích zprava zešíkmené, použil jsem logaritmickou transformaci pomocí přirozeného logaritmu vynásobeného, pro komfort, deseti. Rozpětí transformované proměnné bylo teoreticky od 0 do 18, prakticky samozřejmě užší.

Popisné statistiky transformované škály udává tabulka III.1. V tabulce je vidět, že v 19 letech došlo ke skokovému úbytku respondentů, protože to byla první vlna výzkumu ELSPAC, kdy o účasti rozhodovali oni sami, namísto jejich rodičů, jak tomu bylo v předchozích vlnách. Z ML odhadů průměru a směrodatné odchylky, které poskytl Mplus v rámci FIML odhadu nepodmíněného LGC modelu, je zřejmé, že samovýběr v 19 letech mírně zvyšoval pravděpodobnost účasti těch, kdo v minulosti užívali návykové látky méně než jejich vrstevníci. Tento efekt však není příliš velký. Korelace mezi vlnami jsou středně vysoké a se vzdáleností věků očekávatelně klesají. V reportovaném užívání se tak zdá být určitá míra stability, která by mohla představovat individuální vývoj. V souladu s očekáváním je i postupný nárůst míry užívání, který je nejprudší mezi 15. a 17. rokem.

S ohledem na množství chybějících dat vychází následující analýzy ze vzorku omezeného na ty, kdo se zúčastnili alespoň dvou sběrů dat ($N = 555$). I když takové omezení není vždy nutné a informace od jednorázových účastníků se v modelu využije, může to být spojeno s nestabilitou odhadů a konvergenčními obtížemi v pozdějších fázích analýzy.

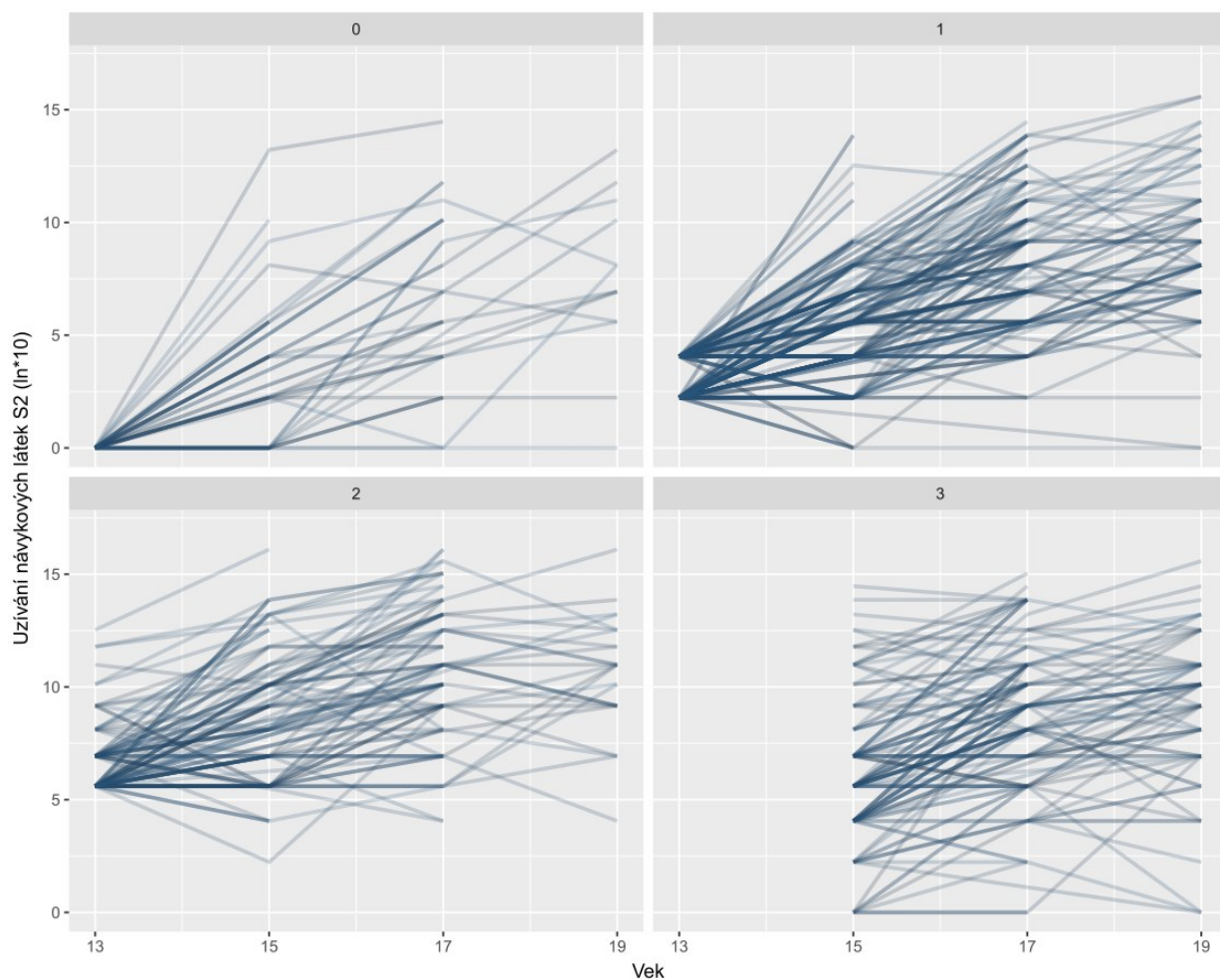
Tabulka III.1. Popisné statistiky transformované škály užívání návykových látek

	N	Min	Max	M	SD	Zešikmení	Strmost	ML odhady			Korelace		
								M	SD	13	15	17	
S2_13ln	576	0,00	13,22	3,77	2,41	0,57	0,66	3,81	2,35				
S2_15ln	511	0,00	16,09	6,01	3,12	0,42	0,21	6,08	3,18	0,55			
S2_17ln	459	0,00	16,09	8,75	3,29	-0,20	-0,33	8,74	3,34	0,43	0,67		
S2_19ln	188	0,00	16,09	9,08	3,22	-0,57	0,65	9,61	3,33	0,30	0,52	0,74	

Poznámka. ML odhady M, SD a korelací byly vypočítány Mplus v rámci odhadu nepodmíněného LGCM modelu používajícího FIML odhad.

Individuální růstové křivky prezentované v grafu na obrázku III.1 ovšem poskytují podstatně méně uspořádaný obraz než výše uvedené popisné statistiky. Pro přehlednost zobrazení byl soubor rozdělen do 4 skupin podle hodnoty ve 13 letech. Individuální křivky jsou transparentní, takže sytější odstín reprezentuje překrývající se křivky. V grafu je vidět, že většina křivek je rostoucích a jen málo jich klesá pod hodnoty ze 13 let. Přesto je paleta pozorovaných vývoju velmi široká a zdá se být smysluplné snažit se této rozmanitosti porozumět.

Obrázek III.1. Individuální pozorované růstové křivky škály užívání návykových látek rozdělené podle hodnoty škály ve 13 letech.



Krok 1 – Specifikace nepodmíněného modelu růstových křivek

V datech jsou k dispozici pouze čtyři měření, což klade limity na to, jak složitý může být model vývoje. Výchozím modelem růstových křivek je zde lineární latentní růstový model s náhodnými průsečíky a směrnici. Ten stanovuje, že rizikové chování lineárně roste s věkem, přičemž výchozí míra rizikového chování i tempo nárůstu (či úbytku) jsou pro každého jednotlivce jiné. Tyto odlišnosti mezi jednotlivci ve výchozí míře (průsečík) a tempu růstu mají normální rozložení, jejichž průměry a rozptyly se odhadují spolu s ostatními parametry modelu. Odhaduje se také kovariance mezi průsečíky a směrnici. Tento výchozí model (označený *lgcm_01*) má jednoduchou reziduální strukturu, která je výchozí strukturou v Mplus – v každém čase tedy může být reziduální rozptyl jiný a rezidua napříč měřeními nekorelují. Model *lgcm_01* tedy popisuje růst pomocí 9 parametrů – průměr a rozptyl počáteční míry užívání (η_1), průměr a rozptyl tempa růstu (η_2), jejich kovariance a 4 reziduální rozptyly. Model byl stejně jako všechny ostatní v této kapitole odhadnut pomocí software Mplus verze 8 (Muthén & Muthén, 2017). Ukazatele shody tohoto modelu s daty jsou spolu se statistikami dalších modelů uvedeny v tabulce III.2. Model *lgcm_01* nevykazuje dobrou shodu s daty, a to prizmatem všech ukazatelů – CFI a TLI jsou při 0,8 příliš nízké, RMSEA a SRMR vysoce překračují hodnotu 0,1, což je také obtížně akceptovatelné. Pro tento model jsou odhadnuté hodnoty průměrné počáteční míry konzumace 3,9 (SD = 2,1) průměrného nárůstu 1,1 (SD = 0,5) za rok. Korelace mezi počáteční mírou a tempem nárůstu byla odhadnuta jako minimální (-0,12 při SE = 0,14). Reziduální rozptyly se pohybovaly mezi 1,4 a 4,2, což je poměrně velké rozpětí.

Model *lgcm_02* oproti modelu *lgcm_01* přidává požadavek stejných reziduí napříč vlnami měření. I když má kvůli tomuto omezení model *lgcm_02* mírně horší shodu s daty než *lgcm_01*, není rozdíl příliš velký a prizmatem ukazatelů, které penalizují za počet parametrů, je *lgcm_02* dokonce lepším modelem (TLI, BIC, RMSEA). Lze tedy říci, že rozdílnost reziduálních rozptylů napříč vlnami může být vysvětlitelná náhodou. Shoda modelu s daty však zůstává špatná. Parametry růstu se od výchozího modelu liší jen málo. Střední hodnoty počáteční míry a tempa růstu (tj. parametry střední růstové křivky) jsou identické jako ve výchozím modelu – 3,9 a 1,1. Co mírně pokleslo, je variabilita těchto náhodných koeficientů – směrodatná odchylka počáteční míry klesla na 1,8 a SD tempa růstu na 0,4. Narostla naopak korelace mezi počáteční mírou a tempem růstu, a to na hodnotu 0,30 (SE = 0,17). Reziduální rozptyl ve všech časech měření byl 3,3. Dalo by se tedy říci, že fixování rozptylu na 0 zde vedlo k mírnému přesunu variability z růstové části modelu do reziduální části modelu. Kdyby byl cílem analýzy samotný LGC model, byl by *lgcm_02* asi přijatelnější. Vzhledem k tomu, že cílem je analýza heterogenity růstových křivek, je přesun této variability do reziduálního rozptylu, kde již nebude dále analyzována, nežádoucí.

Model *lgcm_03* je opět variantou modelu *lgcm_01* s jinou reziduální strukturou. Ta zde má podobu autokorelovaných po sobě následujících reziduí (*lag-1 autocorrelation*), což odpovídá představě, že postupně se měnící rizikové chování není jedinou nenáhodnou složkou jednotlivých měření. Hodnoty rizikového chování takto v modelu s autokorelovanými rezidui korelují nejen díky stabilně se vyvíjejícímu rizikovému chování jako rysu, ale také díky stabilitě prostředí či dalších vlivů, které způsobují to, že sousedící rezidua spolu korelují. Bez omezení parametrů tento model konverguje k nepřijatelnému řešení, kdy právě matice reziduálních rozptylů a kovariancí není pozitivně definitní. Model *lgcm_03a* fixuje všechny tři autokorelace na stejnou hodnotu, model *lgcm_03b* fixuje reziduální rozptyly na stejnou hodnotu a model *lgcm_03c* fixuje jak reziduální rozptyly, tak jejich

autokorelace. Teprve třetí z nich znamená zlepšení ukazatelů shody modelu s daty, byť ty stále nejsou uspokojivé. Ovšem i tento model konverguje k nepřijatelnému řešení, přičemž problém je v matici psí, tedy ve rozptylech a kovarianci počáteční míry a tempa růstu. Pouze *lgcm_03a* konverguje k přijatelnému řešení, jehož shodu s daty nelze označit za zlepšení oproti výchozímu modelu. Pomínu-li problémy s odhadem tohoto modelu, dílčí výsledky ukazují, že ještě větší část variability růstových křivek se přesouvá do komplexní reziduální struktury.

Tabulka III.2. Nepodmíněné modely latentních růstových křivek.

Model	N par	Chi2	DF	p	LL	CFI	TLI	AIC	BIC	SABIC	RMSEA	SRMR
<i>lgcm_01</i>	9	73,9	5	<0,001	-3350,5	0,814	0,777	6719,0	6757,8	6729,3	0,158	0,138
<i>lgcm_02</i>	6	83,1	8	<0,001	-3359,0	0,797	0,848	6730,0	6755,9	6736,8	0,130	0,189
<i>lgcm_03*</i>	12	58,3	2	<0,001	-3341,9	0,848	0,544	6707,8	6759,6	6721,5	0,225	0,071
<i>lgcm_03a</i>	10	66,5	4	<0,001	-3347,3	0,831	0,747	6714,7	6757,9	6726,1	0,168	0,109
<i>lgcm_03b</i>	9	75,3	5	<0,001	-3352,6	0,810	0,772	6723,2	6762,1	6733,5	0,159	0,140
<i>lgcm_03c</i>	7	71,0	7	<0,001	-3352,9	0,827	0,852	6719,8	6750,0	6727,8	0,128	0,137
<i>lgcm_04</i>	11	18,6	3	<0,001	-3323,5	0,958	0,916	6669,0	6716,5	6681,6	0,097	0,105
<i>lgcm_05</i>	8	38,4	6	<0,001	-3335,0	0,912	0,912	6686,0	6720,6	6695,2	0,099	0,152
<i>lgcm_05a</i>	11	18,6	3	<0,001	-3323,5	0,958	0,916	6669,0	6716,5	6681,6	0,097	0,105
<i>lgcm_06</i>	9	32,0	5	<0,001	-3335,0	0,927	0,912	6688,0	6726,9	6698,3	0,099	0,152

Poznámka. *nepřijatelné řešení. N par – počet parametrů

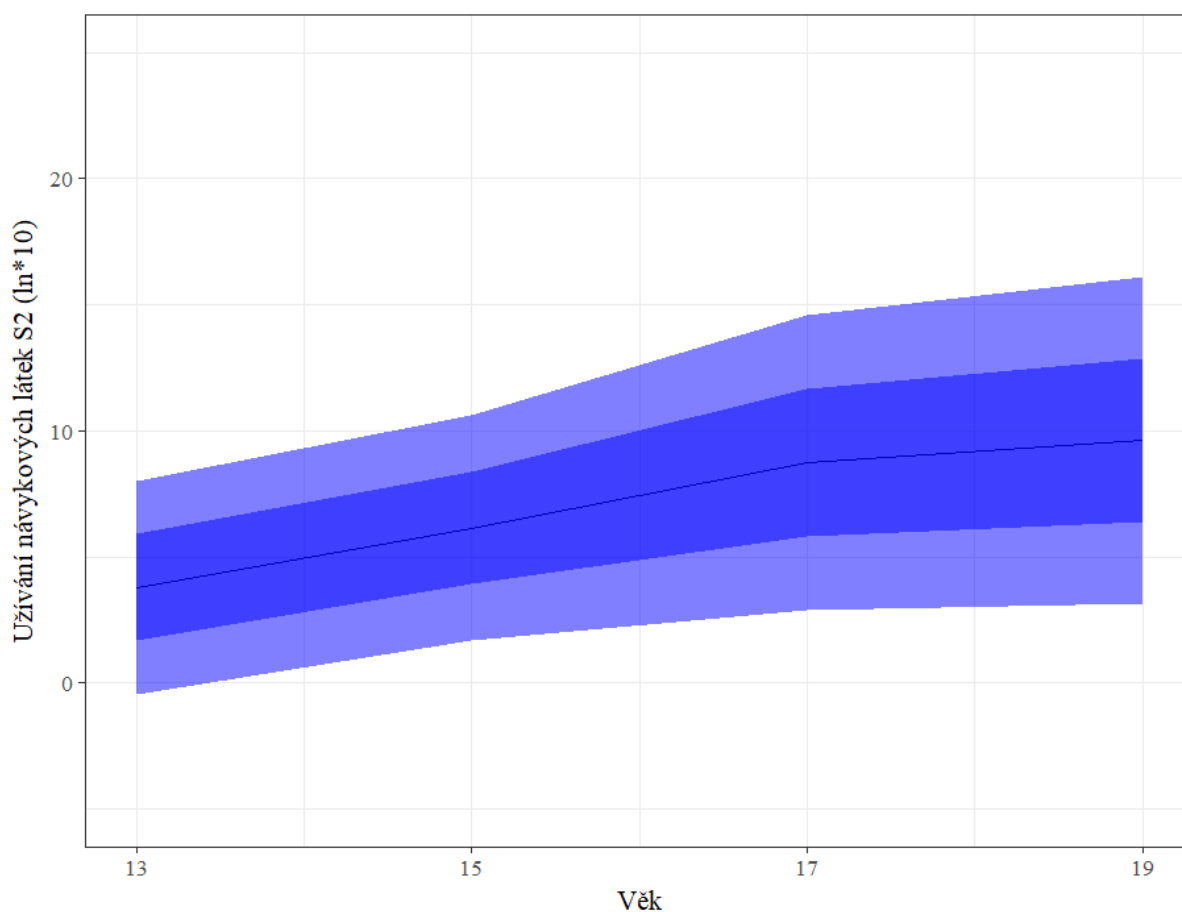
V tomto bodě analýzy se zdá zřejmé, že elaborací reziduální struktury lineárního růstového modelu nelze dosáhnout potřebné shody modelu s daty. Není to překvapivé, protože individuální pozorované růstové křivky jsou v mnoha případech nelineární. Vzhledem k pouhým čtyřem měřením se však z individuálních křivek na nelinearitu vývoje usuzuje obtížně. Nelineární vývoj míry užívání návykových látek dává smysl i teoreticky, když v období dospívání nastává několik příležitostí k započatí pravidelnějšího užívání – prodej cigaret od 16, alkoholu od 18 i neformální sociální aktivity v kultuře s vysokou tolerancí k užívání alkoholu a cigaret.

Následující rozvinutí výchozího modelu je **latent-base model**, tedy model, v němž je tvar růstové křivky odhadován mezi pevně stanoveným počátečním a koncovým bodem. Oproti výchozímu modelu jsou pouze uvolněny časové báze 2. a 3. měření (15 let, 17 let). Místo hodnot 2 a 4, které měly pevně nastavené v lineárním modelu jim byly odhadnuty hodnoty 2,4 a 5,1 (viz obrázek III.2). To znamená, že růst je oproti lineárnímu mírně rychlejší. Tyto hodnoty jsou odhadovány jako fixované, mají tedy v modelu stejnou hodnotu pro všechny účastníky – základní tvar růstové křivky je pro všechny stejný. Jednotlivci se v tomto modelu mohou lišit v počáteční hodnotě růstové křivky ($M=3,79$, $SD = 2,12$) a v tempu růstu (které křivku naklání, $M= 0,97$, $SD = 0,48$), které spolu pouze minimálně korelují ($r_{15} = 0,17$). Není triviální si představit takové rozložení růstových křivek, a proto jej uvádím na obrázku III.3.

Je na něm vidět, že model dobře zachycuje jak nelinearitu vývoje, tak postupný nárůst rozptylu užívání návykových látek až do 17 let a jeho následný pokles. I díky tomu vykazuje model lepší ukazatele shody s daty, byť hodnoty RMSEA a SRMR těsně kolem 0,1 nejsou v absolutním smyslu přijatelné. Zde už však je možné, že nedostatečnou shodu má na svědomí heterogenita růstových křivek, kterou nelze jednoduše popsat normálním rozložením

růstových parametrů, což by měl zajistit teprve mixture model. Za vysokými hodnotami RMSEA a SRMR stojí také jednoduchá reziduální struktura v tomto modelu – rezidua se mohla v jednotlivých věcích lišit (a jejich rozptyly byly odhadnuty na hodnoty od 1,2 po 4,3), ale byla specifikována jako nekorelovaná. Model však svými 11 odhadovanými parametry téměř vyčerpává stupně volnosti ($df=3$), a tak se bez jasné teorie nezdá být vhodné korelace reziduí zařadit. Stojí za to také poznamenat, že modely s málo stupni volnosti (zvláště pak růstové modely s nemnoho měřeními) mají tendenci vykazovat vysoké hodnoty RMSEA, i když model odpovídá datům dobře (Kenny, Kaniskan, & McCoach, 2015). Latent-base model se tak jeví vcelku vhodným kandidátem pro exploraci heterogenity růstových křivek.

Obrázek III.2. Rozložení růstových křivek užívání návykových látek v latent-base modelu (*lgcm_04*). Graf znázorňuje průměrnou růstovou trajektorii a pásma $+1$ SD a $+2$ SD.



Pozorovaná průměrná vývojová trajektorie má esovitý tvar, a tak je smysluplné se pokusit ji modelovat jako logistický růst podle (Grimm & Ram, 2009). Modely *lgcm_05* a *lgcm_05a* popisují individuální trajektorie pomocí logistických křivek s parametry lambda (střed logistické křivky) a alfa (strmost logistické křivky v jejím středu) odhadnutými pro celý vzorek. Rozdíly mezi jednotlivci (jejich křivkami) jsou modelovány pomocí dvou parametrů – počáteční hodnoty růstové křivky ve 13 letech a rychlosti růstu, který v tomto modelu vlastně reprezentuje individuální horní asymptotu růstu (které však v pozorovaném časovém rozpětí nemusí jedinec dosáhnout). Modely 5 a 5a se liší volností odhadu reziduální struktury. Zatímco model 5 má všechna rezidua specifikována jako stejná, v modelu 5a se mohou v jednotlivých věcích lišit. Je zajímavé, že model 5a se s daty shoduje zcela stejně jako model 4. Můžeme tedy hypotetizovat, že růst má logistický tvar. V případě většího množství měření

by byl logistický model parametricky úspornější, ovšem při pouhých 4 měřeních mají latent-base a logistický model parametr stejně. Predikují také stejné křivky, pouze jinak parametrizované. V modelu 5a byl parametr lambda odhadnut na hodnotu 2,15 a alfa na hodnotu 0,90. Počáteční úroveň má v průměru hodnotu 2,9 (SD = 2,2) a tempo růstu 6,9 (SD = 3,4). Hodnota průměrné trajektorie v čase $t = (\text{Věk}-13)$ je tak $2,9 + 6,9(1/(1 + \exp(-(t-2,15)0,90)))$, např. pro věk 13 let tedy $t = 0$ je to 3,78, což je totéž, co průměrná počáteční hodnota v latent-base modelu. Oproti latent-base modelu nám parametr lambda umožňuje si uvědomit, že až do $13+2,15$, tedy 15,15 let rychlost růstu konzumace návykových látek roste a od tohoto věku dále zase klesá. Jinak je ale latent-base model na interpretaci snazší, a tak bych jej pro následné mixture modely použil raději než logistický růst.

Posledním nelineárním modelem je Richardsův růst podle (Grimm & Ram, 2009), který obohacuje logistický růst o jeden parametr navíc. Ten umožňuje logistické růstové křivce nebýt symetrická. V tomto konkrétním případě přidání tohoto parametru do modelu (*lgcm_06*) shodu modelu s daty oproti modelu *lgcm_05* nezvýšilo.

Pro následnou exploraci heterogenity růstových křivek pomocí mixture růstového modelu jsem vybral latent-base model (*lgcm_04*), který dobře popisuje data a je snadno interpretovatelný.

Krok 2 – Modely latentních růstových tříd na latent-base modelu

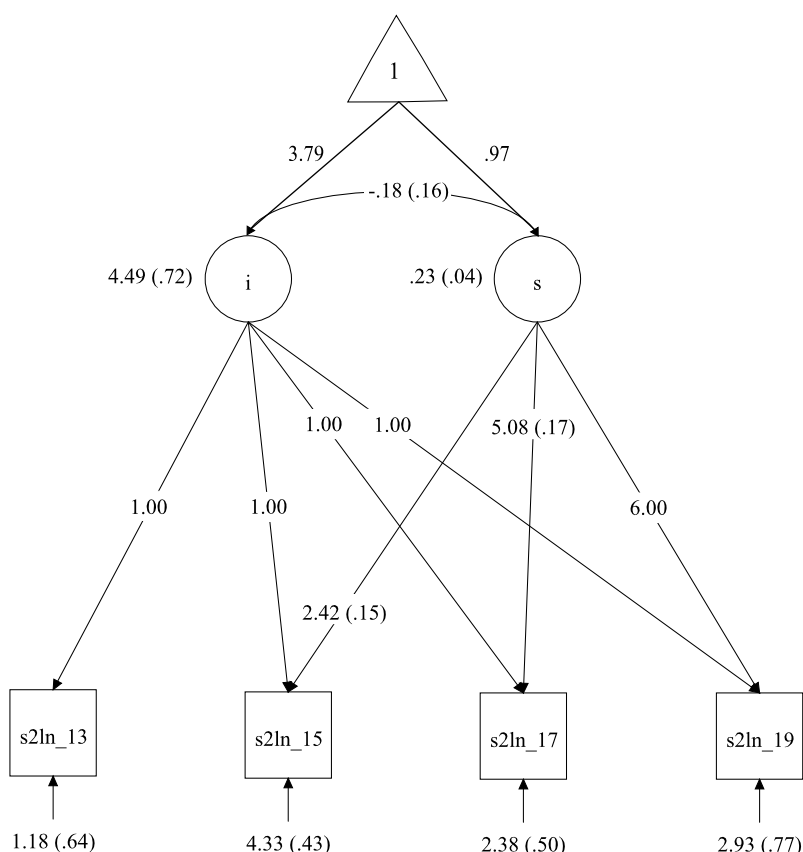
V této části analýzy je cílem zjistit, zda je smysluplné uvažovat o rozdílnosti individuálních růstových křivek tak, že bychom předpokládali existenci určitého počtu latentních tříd, jejichž členové by měli všichni tutéž růstovou křivku až na chybu měření. Růstové křivky různých latentních tříd se pochopitelně liší. Pro referenci je výchozí latent-base znázorněn na obrázku III.3 a můžeme o něm uvažovat jako o LCG modelu s jednou třídou. Pro modely latentních růstových tříd je definující to, že průměrné růstové parametry – zde I a S – jsou v rámci tříd konstantní a mezi třídami se liší. To znamená, že mají uvnitř tříd nulový rozptyl. Ostatní parametry modelu – zde růstové báze a reziduální rozptyly se mezi třídami lišit mohou a nemusí. V následující analýze prezentuji dva druhy modelů latentních růstových tříd. V těch prvních se třídy liší pouze průměry růstových parametrů I a S . V těch prvních tak mají růstové křivky ve všech třídách stejný tvar a počet odhadovaných parametrů neroste s třídami tak rychle. V těch druhých mohou mít díky různým růstovým bázím latentní třídy různý tvar růstové trajektorie, což ovšem znamená další dva parametry s každou třídou navíc. Oba typy modelů předpokládají stejná rezidua napříč třídami. Jejich uvolnění by znamenalo další 4 parametry na každou třídu, a proto by mělo smysl pouze na základě explicitních teoretických předpokladů.

LCGA MODELY SE STEJNÝMI RŮSTOVÝMI BÁZEMI NAPŘÍČ TŘÍDAMI

Postupně jsem odhadl modely s jednou až sedmi latentními třídami. Jejich souhrnné ukazatele jsou uvedeny v tabulce III.3. Na prvním řádku tabulky jsou pro srovnání uvedeny ukazatele shody modelu s daty výchozího LGCM latent-base modelu. Ve druhém sloupci je uveden počet parametrů každého modelu. Každá třída navíc zde znamená další tři parametry – dva průměry růstových parametrů I a S a relativní četnost třídy. Model s jednou třídou ukazuje, oč by byla shoda modelu s daty horší, kdybychom předpokládali, že všichni mají tutéž růstovou křivku. Již u modelu se dvěma latentními třídami se počet parametrů dostává na stejnou hodnotu jako u LGCM modelu, a přitom LCG model nepopisuje data tak dobře.

Prizmatem BIC, který penalizuje počet parametrů modelu dokonce žádný z modelů latentních tříd nereprezentuje data tak věrně jako LGCM model. Odhlédneme-li od parsimonie, až model se 5 třídami nabízí vyšší shodu s daty než LGCM model. Parametrický popis rozložení růstových křivek (ve smyslu popisu jejich rozložení jako normálního rozložení se dvěma parametry) se v tomto případě jeví být vhodnější, než neparametrický prostřednictvím množství diskretních tříd. Tomu nasvědčují i nepřesvědčivé hodnoty entropie, které od hodnoty 0,72 u modelu se třemi třídami už jen postupně klesají k hodnotě 0,62.

Obrázek III.3. Latent-base růstový model užívání návykových látek



Tabulka III.3. Souhrnné ukazatele modelů latentních růstových tříd užívání návykových látek založených na latent-base modelu. Na prvním řádku pro srovnání výchozí LGC model.

Počet tříd	par	LL	AIC	BIC	aBIC	Entropie	LMR adj	LMR p	BLRT	BLRT p	n draws
LGCM	11	-3324	6669	6717	6682						
1	8	-3522	7060	7094	7069						
2	11	-3410	6842	6890	6855	0,604					
3	14	-3348	6724	6784	6740	0,719	117,9	<0,001	124,1	<0,001	5
4	17	-3332	6698	6771	6717	0,657	30,4	0,092	32,0	<0,001	20
5	20	-3322	6683	6770	6706	0,685	19,7	0,145	20,7	<0,001	20
6	23	-3312	6670	6769	6696	0,619	18,6	0,400	19,6	<0,001	20
7	26	-3307	6665	6778	6695	0,617	9,9	0,330	10,4	0,05	100

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. N draws – počet bootstrapových vzorků použitých algoritmem.

Pro volbu optimálního modelu latentních tříd je třeba přihlídnout také k odhadnutým parametrům jednotlivých modelů. Ty jsou uvedeny v tabulce III.4. Pro snadnější interpretaci obsahuje obrázek III.4 grafy růstových křivek odpovídajících jednotlivým latentním třídám.

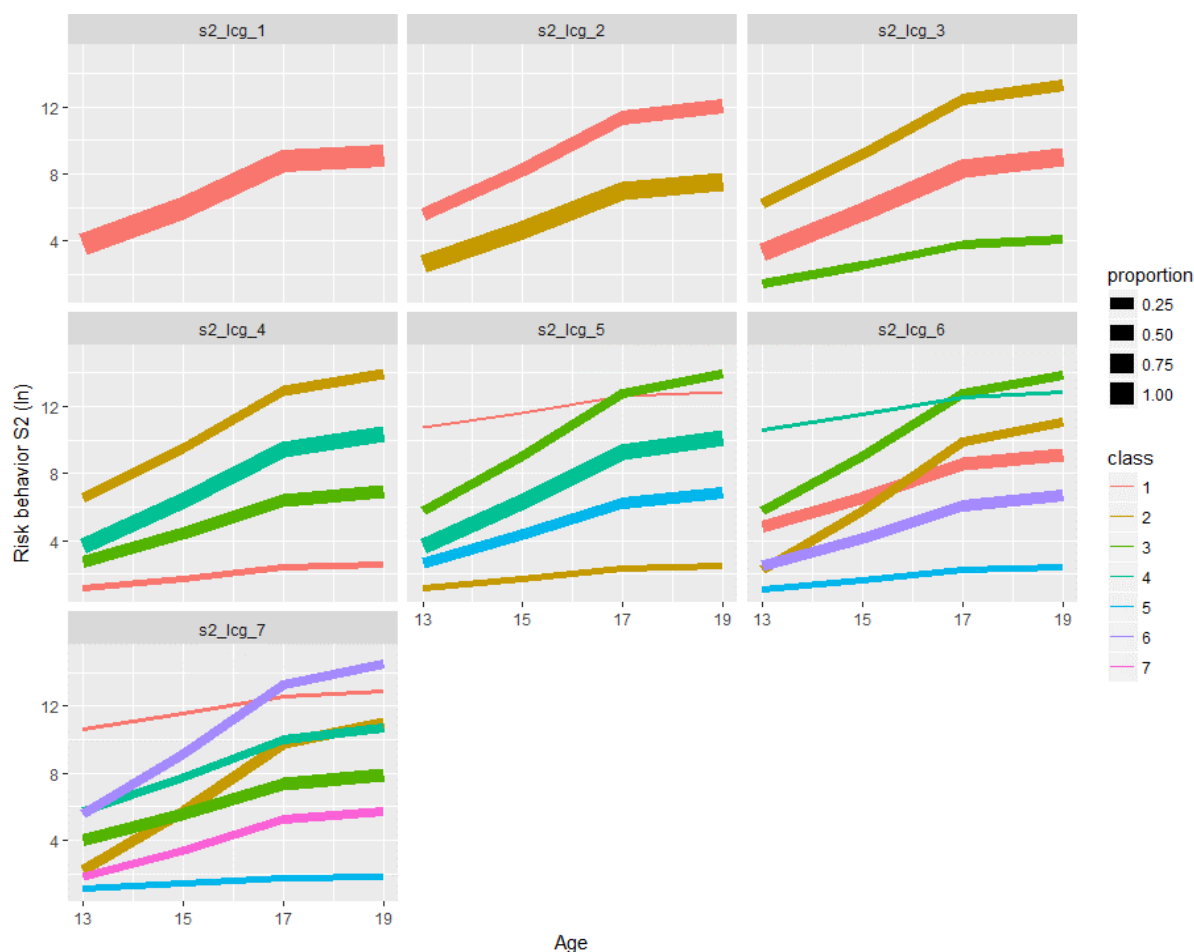
Tabulka III.4. Modely latentních růstových tříd užívání návykových látek s rezidui a časovými bázemi konstantními napříč třídami.

Parametr	C	Model s počtem tříd...														
		1	SE	2	SE	3	SE	4	SE	5	SE	6	SE	7	SE	
Relativní četnosti	1	1		0,38		0,65		0,06		0,01		0,29		0,02		
tříd	2			0,62		0,23		0,19		0,06		0,20		0,24		
	3					0,12		0,28		0,19		0,19		0,29		
	4							0,48		0,48		0,02		0,11		
	5									0,26		0,05		0,04		
	6											0,26		0,16		
	7													0,14		
Průměry	I	1	3,82	0,12	5,60	0,44	3,32	0,19	1,18	0,30	10,80	0,95	4,84	0,52	10,63	0,96
růstových	S	1	0,88	0,04	1,08	0,06	0,95	0,05	0,23	0,17	0,35	0,33	0,72	0,10	0,38	0,30
parametrů	I	2			2,61	0,21	6,24	0,39	6,52	0,51	1,14	0,30	2,22	0,31	2,16	0,28
	S	2			0,82	0,06	1,17	0,07	1,23	0,10	0,23	0,17	1,48	0,10	1,47	0,08
	I	3					1,43	0,29	2,75	0,36	5,80	0,34	5,75	0,44	3,97	0,87
	S	3					0,45	0,10	0,70	0,07	1,35	0,10	1,36	0,13	0,65	0,08
	I	4							3,67	0,23	3,66	0,23	10,63	0,97	5,71	0,49
	S	4							1,12	0,08	1,08	0,07	0,38	0,28	0,83	0,17
	I	5									2,65	0,38	1,09	0,33	1,11	0,43
	S	5									0,70	0,06	0,22	0,20	0,11	0,21
	I	6											2,49	0,70	5,55	0,44
	S	6											0,70	0,07	1,49	0,14
	I	7													1,83	0,91
	S	7													0,65	0,08
Reziduální rozptyly	13	1	5,55	0,44	3,71	0,36	3,61	0,34	3,80	0,34	3,36	0,30	2,38	0,42	2,28	0,27
	15	1	10,21	0,74	6,14	0,47	4,72	0,42	4,72	0,45	4,51	0,44	4,46	0,43	4,31	0,40
	17	1	10,86	0,70	6,30	0,51	4,11	0,38	2,96	0,47	2,92	0,48	2,80	0,51	2,13	0,79
	19	1	10,39	1,22	7,06	1,04	5,06	0,60	3,64	0,66	3,66	0,73	3,40	0,79	3,77	1,11
Časové báze (Věk-13)	13	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	15	1	2,47	0,19	2,49	0,17	2,52	0,16	2,44	0,15	2,42	0,15	2,40	0,16	2,41	0,15
	17	1	5,66	0,25	5,36	0,20	5,29	0,18	5,17	0,16	5,19	0,16	5,17	0,17	5,19	0,17
	19	1	6,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Čas je kódovaný jako Věk – 13, časové báze pro věk 15 a 17 jsou volně odhadované.

Volba optimálního modelu zde není snadná, protože jednotlivá kritéria vedou k různým modelům. Informační kritéria v čele v BIC zde upřednostňují modely s více latentními třídami – klesají až k modelu se 7 třídami, BIC je nejnižší u modelu se 6 třídami. LRT testy se výrazně liší. Podle LMR testu je model se 3 třídami lepší než model se 2 třídami, ale vyšší počty tříd již podle tohoto testu nevedou k signifikantně lépe pasujícím modelům. Naopak bootstrapový LRT nalézá signifikantní nárůst shody modelu s daty až do modelu se 6 třídami a na tento verdikt stačilo *optional-stopping* algoritmu pouhých 20 bootstrapových vzorků. Entropie dosahuje nejvyšší hodnoty již u modelu se 3 třídami, což naznačuje spíše spojitost rozložení růstových trajektorií. Až do modelu se 4 třídami mají všechny třídy relativní četnosti vyšší než 5 %, ale v modelu s 5 třídami je jedna třída odhadnuta na pouhé 1 % populace, s čímž se poji zvýšené riziko toho, že jde o nereplikovatelný artefakt zpracovávaných dat. V dalších modelech už se pak tak malá třída nevyskytuje. Pohledem na růstové trajektorie jednotlivých tříd je vidět, jak se s rostoucím počtem tříd vytvářejí trajektorie „vějíř“ pokrývající rozpětí pozorovaných individuálních trajektorií s extrémnějšími třídami méně četnými, než jsou třídy popisující stření vývoj. Od modelu s 5 třídami dále se k tomuto vzorci přidává velmi malá třída s konstantně vysokými hodnotami. V modelu se 6 třídami se střední dvě třídy štěpí každá na dvě třídy – jednu s rychlejším růstem a jednu s pomalejším růstem. Konečně, od modelu se 4 třídami mírně klesá časová báze pro pozorování v 17 letech, čímž se predikované trajektorie v daném modelu nepatrně napřimují.

Obrázek III.4. Růstové trajektorie jednotlivých latentních tříd v modelech s 1 až 7 latentními třídami.



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

Z hlediska interpretovatelnosti modelu bychom mohli považovat za optimální model se 6 třídami, na který také ukazují dva nejdoporučovanější ukazatele – BIC a BLR test. Nevýhodou je poměrně nízká hodnota entropie, což značí, že klasifikace jednotlivců do tříd je spojena se značnou mírou nejistoty. To komplikuje případné hledání kovariátů členství ve třídě. Očekávali bychom pak, že v populaci je malá část dospívajících, kteří se po celou dobu dospívání návykovým látkám vyhýbají, a ještě menší část, která je naopak stabilně užívá ve vysoké míře. Nejčastější, mírné užívání ve 13 letech může pokračovat mírným, nebo rychlejším růstem užívání. Žádná z odhadnutých tříd nemá trajektorii klesající, nebo stagnující. Takoví účastníci v datech nejsou, což může znamenat nereprezentativnost vzorku, nebo to, že se takové vývojové trajektorie v populaci vyskytují velmi zřídka či vůbec.

LCGA MODELY S RŮZNÝMI RŮSTOVÝMI BÁZEMI NAPŘÍČ TŘÍDAMI

Postupně jsem odhadl modely s jednou až pěti latentními třídami, které se mohly lišit i tvarem své trajektorie, protože časové báze pro věk 15 a 17 se volně odhadovaly. Jejich souhrnné ukazatele jsou uvedeny v tabulce III.5. Na prvním řádku tabulky jsou pro srovnání uvedeny ukazatele shody modelu s daty výchozího LGC latent-base modelu. Ve druhém sloupci je uveden počet parametrů každého modelu. Každá třída navíc zde znamená 5 dalších parametrů – relativní četnost třídy, průměrný průsečík a směrnice a dvě časové báze. Počet parametrů tak s počtem tříd narůstá velmi rychle. Stejně jako u předchozích modelů, tabulka III.6 uvádí jednotlivé parametry modelů a obrázek III.5 grafy růstových křivek náležitých latentním třídám ve všech 5 modelech.

Tabulka III.5. Souhrnné ukazatele modelů latentních růstových tříd užívání návykových látek založených na latent-base modelu. Na prvním řádku pro srovnání výchozí LGC model.

Počet tříd	par	LL	AIC	BIC	aBIC	Entropie	LMR adj	LMR p	BLRT	BLRT p	n draws
LGCM	11	-3324	6669	6717	6682						
1	8	-3522	7060	7094	7069						
2	13	-3404	6833	6890	6848	0,636	-	-	-	-	-
3	18	-3334	6703	6781	6724	0,731	135,8	0,010	140,1	<0,001	5
4	23	-3313	6671	6771	6698	0,668	41,0	0,485	42,1	<0,001	20
5	28	-3303	6662	6783	6694	0,729	21,8	0,247	21,8	<0,001	50
5plus	22	-3307	6658	6753	6684	0,735	-	-	-	-	-

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální.

Z perspektivy shody modelu s daty nabízí modely s různými růstovými bázemi napříč třídami podobný obrázek jako předchozí modely. Informační kritéria preferují modely s vyšším počtem tříd. Poněkud překvapivě však nyní s rostoucím počtem tříd roste i ukazatel entropie, což je známkou toho, že možnost mít v každé třídě jiný tvar trajektorie je pro stanovení dobře odlišených latentních tříd výhodná. Modely s 3 a 5 mají hodnotu entropie kolem 0,73 a to už je celkem uspokojivá hodnota. Test LMR je opět velmi konzervativní a již model se 4 třídami nepovažuje za statisticky významně lepší než model se 3 třídami. Test BLRT již naráží na komplexitu modelu a pro model s 5 třídami již jeho výpočet po 16 hodinách havaruje. Ovšem model se 4 třídami je podle něj lepší než model s 3 třídami. Všechny modely

až na model se 5 třídami neodhadují žádné skupiny s relativní četností pod 5 %. Model s 5 třídami dvě takové zahrnuje a při dostupné velikosti vzorku se problematičnost ta malých tříd projevuje prudkým nárůstem směrodatných chyb odhadu časovýchází pro tyto dvě třídy (třídy 4 a 5). To je škoda, protože jejich odhadnuté růstové křivky jsou teoreticky plauzibilní.

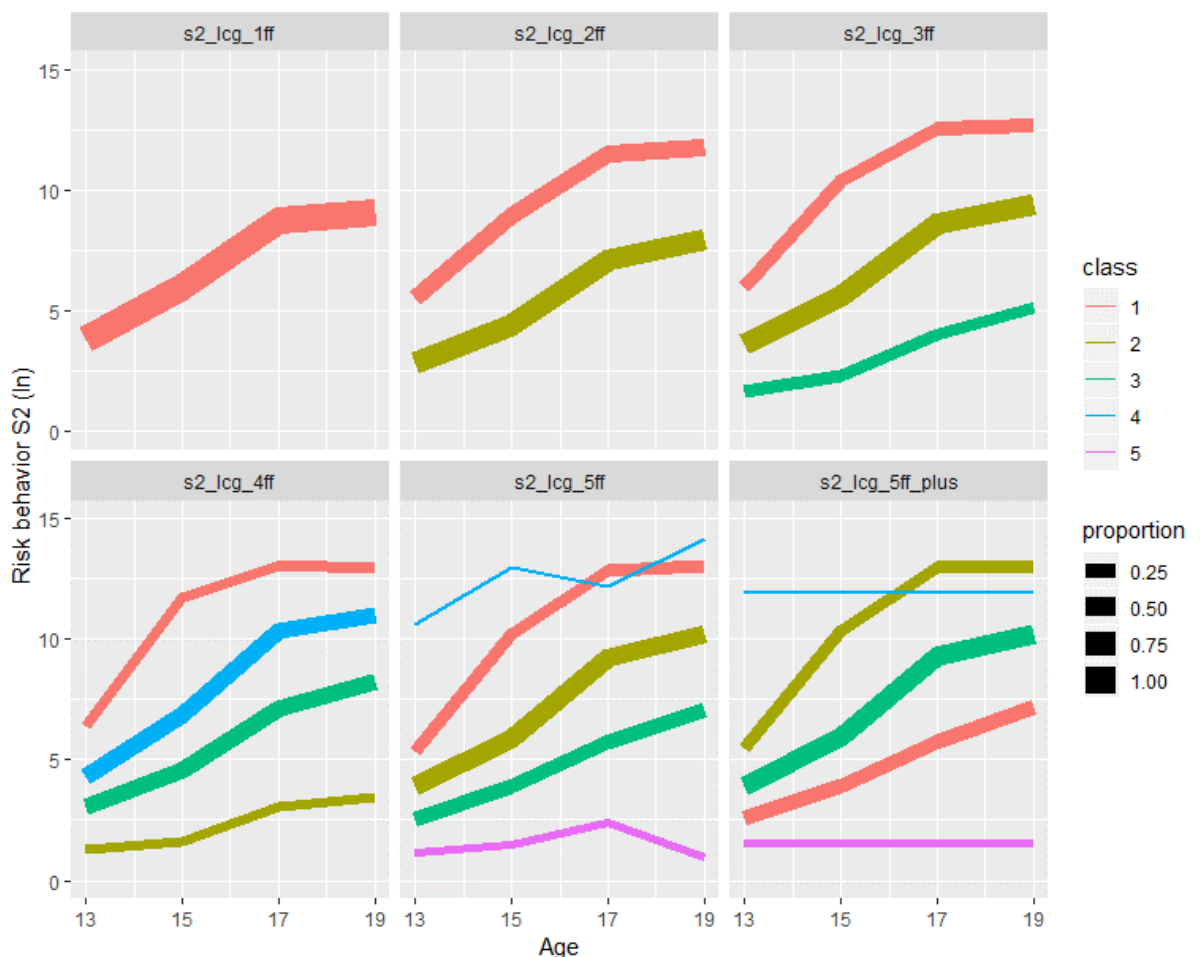
Tabulka III.6. Modely latentních růstových tříd užívání návykových látek s rezidui konstantními napříč třídami a třídě odlišnými časovými bázemi.

		Model s počtem tříd...											
Parametr		C 1	SE	2	SE	3	SE	4	SE	5	SE		
Relativní četnosti tříd		1	1			0,35		0,21		0,12		0,19	
		2				0,65		0,64		0,08		0,50	
		3					0,15		0,42		0,26		
		4							0,38		0,01		
		5									0,04		
Průměry růstových parametrů	I	1	3,82	0,12	5,54	0,36	5,97	0,40	6,42	0,59	5,37	0,44	
	S	1	0,88	0,04	1,04	0,06	1,12	0,09	1,10	0,16	1,28	0,13	
	I	2			2,82	0,21	3,62	0,25	1,29	0,32	3,91	0,23	
	S	2			0,85	0,05	0,96	0,05	0,36	0,42	1,05	0,06	
	I	3					1,69	0,34	3,05	0,53	2,56	0,29	
	S	3					0,57	0,15	0,86	0,14	0,75	0,06	
	I	4							4,29	0,82	10,59	1,26	
	S	4							1,12	0,09	0,60	0,63	
Reziduální rozptyly	I	5									1,15	0,33	
	S	5									-0,03	0,13	
	I	13	1	5,55	0,44	3,83	0,36	3,76	0,35	3,77	0,50	3,51	0,33
	S	15	1	10,21	0,74	5,55	0,42	3,84	0,45	3,05	1,11	3,90	0,40
Časové báze (Věk-13)	I	17	1	10,86	0,70	6,37	0,52	4,22	0,51	3,65	1,18	3,23	0,54
	S	19	1	10,39	1,22	7,32	1,04	5,33	0,66	4,46	1,39	2,80	0,72
	I	15	1	2,47	0,19	3,30	0,39	3,95	0,53	4,83	0,92	3,79	0,45
	S	17	1	5,66	0,25	5,77	0,28	5,84	0,32	6,02	0,48	5,85	0,33
Časové báze (Věk-13)	I	15	2			1,88	0,18	2,04	0,17	0,94	1,21	1,90	0,19
	S	17	2			5,07	0,28	5,19	0,24	4,91	3,05	5,11	0,23
	I	15	3					1,13	0,50	1,76	0,32	1,78	0,36
	S	17	3					4,13	0,83	4,71	0,97	4,24	0,44
Časové báze (Věk-13)	I	15	4							2,30	0,54	4,02	2,87
	S	17	4							5,41	0,29	2,69	3,54
	I	15	5									-12,21	62,15
	S	17	5									-42,79	210,61

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Čas je kódovaný jako Věk – 13, časové báze pro věk 15 a 17 jsou volně odhadované, časové báze pro věky 13 a 19 jsou fixované na 0 a 6.

Tím se dostávám k jádru modelu, a to k růstovým křivkám odhadnutých latentních tříd. Díky uvolnění tvaru jsou křivky podstatně interpretovatelnější a v souladu s doposud popisovanými vývojovými trajektoriemi užívání návykových látek. Model se 3 třídami identifikuje majoritní třídu ($p = 64\%$), v níž dochází k nejvyššímu nárůstu udávaného užívání návykových látek mezi 15. a 17. rokem. Vedle ní pak dvě menší třídy, kdy v jedné ($p = 21\%$) došlo k nejprudšímu nárůstu dříve, mezi 13. a 15. rokem a celková úroveň užívání je vyšší. Brzký start užívání alkoholu a cigaret tak v adolescenci obvykle vede k vyšší míře užívání. Třetí latentní třída ($p = 15\%$) pak popisuje vývoj s nejnižším počátečním stavem užívání a také nejnižším nárůstem. Vějířovité uspořádání trajektorií odpovídá jednoduchému vzorci, čím dříve adolescent začíná užívat a čím více je na počátku adolescence užívá, tím rychlejší je nárůst užívání. Model se 4 třídami tento vzorec jen mírně elaboruje pomyslným⁵ rozdělením střední třídy na dvě – vyšší a nižší. Dosahuje tím zřetelného poklesu reziduí zvláště v 15 a 19 letech. Tyto dvě střední třídy jsou svými trajektoriemi paralelní a dohromady zahrnují 80% populace. Problém je, že model mezi těmito dvěma třídami nedokáže při klasifikaci jednotlivců tak dobře rozlišovat, což má za následek nezanedbatelný pokles entropie.

Obrázek III.5. Růstové trajektorie jednotlivých latentních tříd v modelech s 1 až 5 latentními třídami s rozdílnými časovými bázemi.



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

⁵ Vskutku pomyslným – nejde o hierarchické klastrování, kde by se shluky slučovaly či rozdělovaly.

Model s 5 třídami se zdá být lepší elaborací modelu se 3 třídami, protože jej vlastně „očisťuje“ o extrémní případy, pro které hypotetizuje 4. a 5. třídu – třídy se stabilně velmi vysokým udávaným užíváním a třídy s minimálním užíváním. Obě třídy jsou velmi malé a odhady jejich parametrů zatížené velkou směrodatnou chybou. Vzhledem k tomu, že jediným signifikantním parametrem je v nich průměrná počáteční hodnota, lze je vnímat jako třídy se stabilní úrovní užívání. Proto je pro tyto třídy tak obtížné stanovit časové báze (na zcela vodorovné trajektorii by mohly mít jakoukoli hodnotu a tvar trajektorie by to neovlivnilo). Protože toto vyčlenění extrémních jednotlivců „očistilo“ hlavní tři třídy, které dohromady zahrnují 95 % populace, narostla zpět hodnota entropie. Také rezidua poklesla ve všech věcí vyjma 15. roku. Teoreticky nejnositelnější se tedy i přes statistické neduhy dané nedostatky vzorku jeví být model s 5 třídami.

I když to na první pohled hraničí s rybařením, lze jednotlivé modely s latentními třídami ještě dále ladit. Model s 5 třídami tak lze specifikovat tak, že hledáme 5 tříd, z nichž 3 budou volně odhadované, co se týká růstových faktorů i časových bází, a 2 budou předem specifikované extrémní stabilní třídy – jejich směrnice tedy bude fixována na 0 a časové báze v 15 a 17 letech zafixovány na 0. Odhadnuté parametry tohoto modelu (5plus) jsou velmi podobné výše popsanému modelu s 5 třídami, mají nižší směrodatné chyby a jeho ukazatele shody s daty jsou lepší (BIC = 6753, AIC = 6658, aBIC = 6684). Pokud již bychom byli rozhodnutí o počtu latentních tříd, má smysl model ještě tímto způsobem parametricky zúpsornit a zpřesnit tak odhady parametrů, o které nám primárně jde.

Krok 3 – Růstové mixture modely založené na latent-base modelu

Oproti předchozím modelům latentních tříd uvažuje model směsí tak, že latentní třída nezahrnuje jednotlivce s identickými růstovými křivkami, ale jednotlivce, jejichž růstové křivky, respektive jejich parametry, jsou nějak, obvykle normálně rozložené. Heterogenita uvnitř třídy tedy již nespádá celá do reziduálního rozptylu, ale dělí se na variabilitu individuálních růstových křivek a zbývající reziduální rozptyl (kolem individuálních křivek). Specifikace GMM modelu tedy oproti LCG modelu pouze uvolní rozptyly růstových parametrů uvnitř tříd a jejich kovarianci. GMM model má tak v případě lineárního růstového modelu ještě o tři parametry na třídu více (pokud se nerozhodneme některé z nich zafixovat). Počet odhadovaných parametrů tak roste velmi rychle. Podobně jako u modelů latentních tříd zde budu prezentovat dvě sady modelů vycházející z latent-base LGC modelu – modely, v nichž jsou časové báze napříč třídami totožné, a modely, v nichž je jim umožněno, aby se lišily. Oba typy modelů předpokládají stejná rezidua napříč třídami a pro snížení počtu odhadovaných parametrů také shodná rezidua napříč časy měření.

GMM SE STEJNÝMI RŮSTOVÝMI BÁZEMI NAPŘÍČ TŘÍDAMI

Postupně jsem odhadl modely s jednou až čtyřmi latentními třídami. Modely s více třídami jsem neodhadoval, protože již u modelu se 3 třídami se objevily problémy s odhadem jeho parametrů a ty se u modelu se 4 třídami dále prohloubily. Vzhledem k tomu, že GMM modely část rozdílnosti individuálních růstových křivek přesouvají dovnitř latentních tříd, je očekávatelné, že GMM tříd bude méně. Jejich souhrnné ukazatele jsou uvedeny v tabulce III.7. Na prvním řádku tabulky jsou pro srovnání uvedeny ukazatele shody modelu s daty

výchozího LGC latent-base modelu. Ve druhém sloupci je uveden počet parametrů každého modelu. Každá třída navíc zde znamená dalších 6 parametrů – dva průměry růstových parametrů I a S, jejich rozptyly a kovariance a konečně relativní četnost třídy. GMM model s jednou třídou je vlastně stejný jako LGC model; rozdíl v modelech uvedených v tabulce III.8 spočívá v tom, že v GMM modelu jsou rezidua specifikována jako shodná napříč věky, pokrývá je tedy 1 parametr namísto 4. Ani tato úsporná parametrizace však nezabránila problémům s odhadem parametrů.

Tabulka III.7. Souhrnné ukazatele GMM modelů užívání návykových látek založených na latent-base modelu s konstantními časovými bázemi. Na prvním řádku pro srovnání výchozí LGCM model.

Počet tříd	par	LL	AIC	BIC	aBIC	Entropie	LMR adj	LMR p	BLRT	BLRT p	n draws
LGCM	11	-3324	6669	6717	6682						
1	8	-3335	6686	6721	6695						
2	14	-3326	6680	6740	6696	0,878					
3*	20	-3318	6676	6762	6699	0,776					
4*	26	-3302	6656	6768	6686	0,752					

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. *Nepřípustný odhad modelu.

Postupný odhad modelů s jednou až čtyřmi latentními třídami ukázal, že tímto směrem cesta nevede. Patrně hlavním znakem neuspokojivosti tohoto modelu směsi růstových křivek jsou relativní četnosti extrahovaných latentních tříd: Ve všech modelech od 2 do 4 tříd je jedna velká třída zahrnující přes 90 % populace a zbývající třídy jsou velmi malé. Modely tak v zásadě říkají, že heterogenita růstových křivek je při konstantním tvaru růstové křivky popsána LGC modelem (tedy jednou třídou) tak dobře, že hledání dalších tříd je spíše rybařením, hledáním malých skupinek podobných případů, u nichž je pravděpodobnější, že jde o artefakty vzniku dat než o skutečné velmi malé subpopulace. Protože tyto další třídy jsou velmi malé, algoritmus nemá dostatek dat k určení jejich parametrů a dochází k nepřípustným řešením z důvodu negativních odhadů rozptylu růstových parametrů uvnitř tříd v modelech se 3 a 4 třídami. Jediným přípustným a částečně i smysluplným modelem je zde tedy model se dvěma třídami, který z majority vyděluje malou podskupinu se stabilně vysokým udávaným užíváním návykových látek. Je však otázkou, zda jde o reálné užívání či svého druhu adolescentní vychloubání v dotazníku.

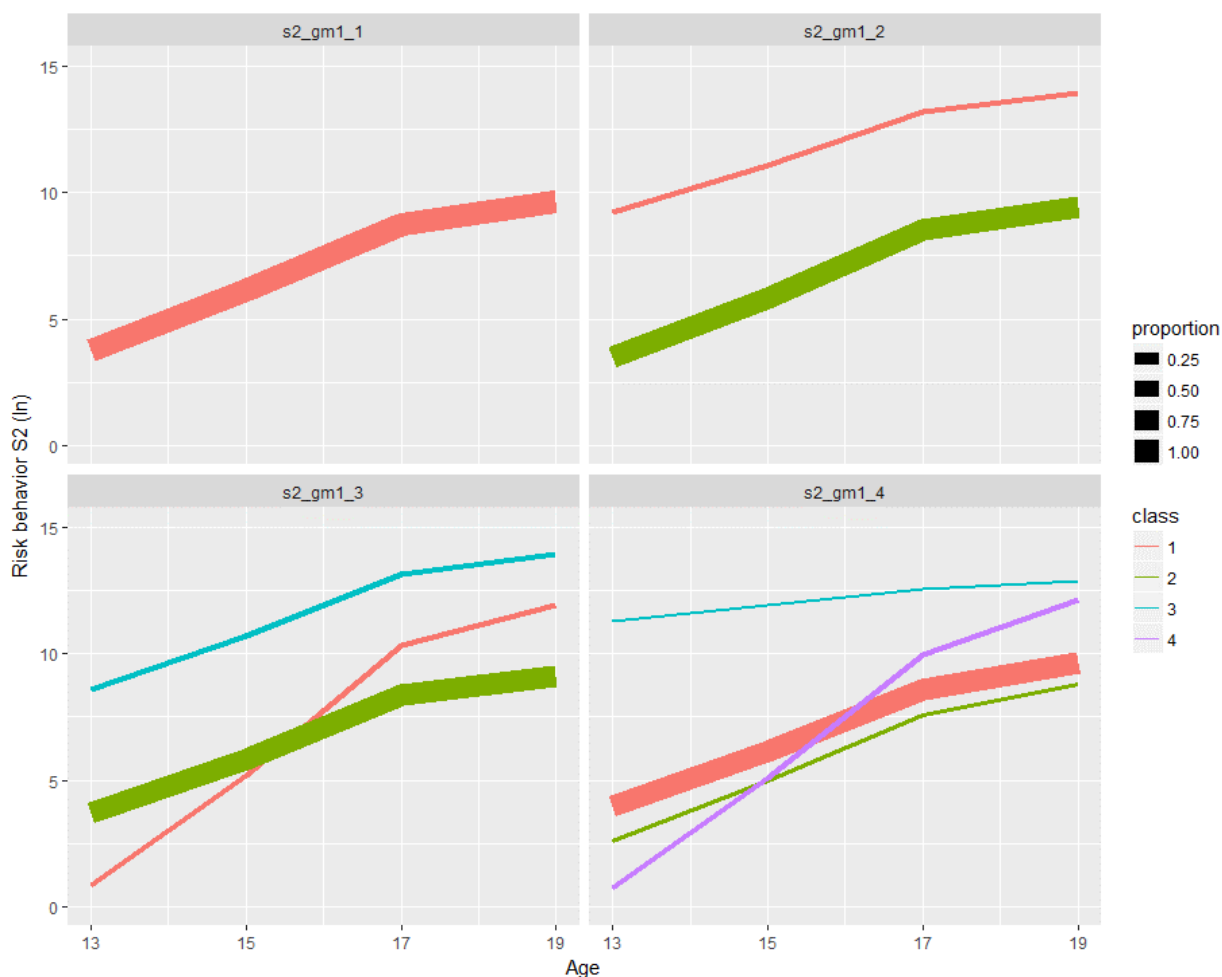
Další cestou, jak model parametricky zúspornit by bylo specifikovat rozptyly růstových parametrů jako shodné napříč latentními třídami. Tím by se zamezilo odhadu negativních rozptylů. Variabilita růstových křivek uvnitř tříd by ta byla ve všech třídách stejná, a to je obtížně akceptovatelný předpoklad, už jen vzhledem k velikosti a extrémnosti tříd. Proto zde tuto modifikaci modelu neprezentují.

Tabulka III.8. GMM modely užívání návykových látek s rezidui a časovými bázemi konstantními napříč třídami.

Parametr		C	1	SE	2	SE	3	SE	4	SE
Relativní četnosti tříd		1	1		0,03		0,03		0,93	
		2			0,97		0,94		0,02	
		3					0,03		0,02	
		4							0,04	
Průměry růstových parametrů	I	1	3,77	0,12	9,20	1,23	0,84	0,46	3,99	0,15
	S	1	0,98	0,04	0,79	0,31	1,85	0,05	0,94	0,04
	I	2			3,52	0,14	3,73	0,16	2,61	0,05
	S	2			0,98	0,04	0,90	0,05	1,03	0,22
	I	3					8,58	2,01	11,29	0,71
	S	3					0,89	0,45	0,26	0,19
	I	4							0,72	0,45
	S	4							1,90	0,04
Rozptyly Růstových parametrů	I	1	3,36	0,55	3,36	2,07	-0,29	0,65	2,33	0,42
	S	1	0,14	0,03	0,28	0,19	-0,13	0,01	0,08	0,03
	I	2			2,10	0,43	1,74	0,41	-2,31	0,17
	S	2			0,14	0,03	0,09	0,04	0,41	0,22
	I	3					5,30	3,45	-0,46	0,97
	S	3					0,35	0,20	0,00	0,17
	I	4							-0,43	0,61
	S	4							-0,13	0,01
Kovariance růstových parametrů	I-S	1	0,17	0,10	-0,87	0,43	0,32	0,06	0,44	0,10
	I-S	2			0,21	0,09	0,39	0,08	0,33	0,07
	I-S	3					-1,24	0,70	0,24	0,49
	I-S	4							0,34	0,05
Reziduální rozptyly	13-19		3,19	0,23	3,15	0,24	3,15	0,23	3,13	0,23
Časové báze (Věk-13)	13	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	15	1	2,41	0,15	2,38	0,16	2,34	0,16	2,29	0,07
	17	1	5,06	0,15	5,08	0,16	5,12	0,16	4,86	0,03
	19	1	6,00	0,00	6,00	0,00	6,00	0,00	6,00	0,00

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Čas je kódovaný jako Věk – 13, časové báze pro věk 15 a 17 jsou volně odhadované.

Obrázek III.6. Růstové trajektorie jednotlivých latentních tříd v GMM modelech s 1 až 4 latentními třídami s konstantními časovými bázemi.



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Rozptyl růstových křivek uvnitř tříd není znázorněn. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

GMM S RŮZNÝMI RŮSTOVÝMI BÁZEMI NAPŘÍČ TŘÍDAMI

Lze očekávat, že stejně jako LCG modelů bude i u GMM modelů vývoje užívání návykových látek reálnější předpokládat, že různé latentní třídy mají různý tvar střední vývojové trajektorie. Modely, které v této části popíšeme, to umožňují. Zároveň umožňují, aby se třídy lišily v rozptylu růstových parametrů. Naopak reziduální struktura je nejjednodušší možná – stejná nekorelovaná rezidua napříč věky i latentními třídami. Postupně jsem odhadl modely s jednou až čtyřmi latentními třídami. Jejich souhrnné ukazatele jsou uvedeny v tabulce III.9. Na prvním řádku tabulky jsou pro srovnání uvedeny ukazatele shody modelu s daty výchozího LGC latent-base modelu. Ve druhém sloupci je uveden počet parametrů každého modelu. Každá třída navíc zde znamená dalších 8 parametrů – dva průměry růstových parametrů I a S, jejich rozptyly a kovariance, dvě časové báze a konečně relativní četnost třídy.

Tabulka III.9. Souhrnné ukazatele GMM modelů užívání návykových látek založených na latent-base modelu s různými časovými bázemi napříč třídami. Na prvním řádku pro srovnání výchozí LGCM model.

Počet tříd	par	LL	AIC	BIC	aBIC	Entropie	LMR adj	LMR p	BLRT	BLRT p
LGCM	11	-3324	6669	6717	6682					
1	8	-3335	6686	6721	6695					
2	16	-3295	6623	6692	6641	0,763	77,7	<0,001	79,2	<0,001
3	24	-3283	6614	6717	6641	0,726	24,7	0,138	25,2	0,86
4*	29	-3273	6604	6730	6637	0,660	21,1	0,265	21,6	0,375

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. *rozptyly směrnice zafixovány na stejnou hodnotu pro dosažení konvergence. Počet bootstrapových vzorků pevně nastaven na 100.

Informační kritéria, která tolik nepenalizují za počet parametrů klesala až k modelu se 4 třídami. Naopak nejkonzervativnější BIC a tentokrát spolu s ním i BLR test a LMR test preferují model se dvěma třídami. U modelu se dvěma třídami je také nejvyšší hodnota entropie. Odhad modelu se 4 třídami již byl z hlediska konvergence poměrně obtížný a podařilo se jej odhadnout jen za cenu nastavení rozptylů směrnic na stejnou hodnotu napříč třídami. Ani to však nebylo dokonalé řešení, protože i když se díky tomu podařilo dosáhnout konvergence, jedna kovariance (ve třídě 3) byla odhadnuta na nepřipustně vysokou hodnotu, která by po převodu na korelaci znamenala absolutní hodnotu vyšší než 1.

Podobně jako u předchozích GMM modelů, mají i tyto tendenci odhadovat jednu velkou třídu doplněnou malými třídami. V tomto případě jsou však velikosti malých tříd přijatelné (nejnižší hodnota jsou 2 %) a jejich růstové parametry v souladu s teorií. V případě modelu se dvěma třídami vychází dvě třídy, které se v předchozích modelech pravidelně objevují, a to je třída „střední“, majoritní, zahrnující 90 % populace, jejíž udávaná míra užívání návykových látek stoupá mírně nelineárně od poměrně nízkých hodnot ke středně vysokým. Variabilita růstových křivek v této třídě je nezanedbatelná, ale ne tak vysoká, aby třída ztrácela svůj charakter. Počáteční hodnoty růstových křivek mají směrodatnou odchylku 1,5 (na logaritmické škále 0–16) a směrnice mají SD = 0,45. Protože kovariance počátečních hodnot a směrnic je minimální, je míře rostoucí variabilita užívání návykových látek v této třídě právě projevem různých temp růstu užívání. Protože průměrné tempo růstu je 0,96, může v extrémních případech do této třídy patřit i někdo, jehož tempo růstu užívání je minimální až nulové. Pozorovanou variabilitu míry užívání, která je rozdílnými vývoji jednotlivců v rámci latentní třídy způsobena, je poměrně obtížné si na základě růstových parametrů představit. Proto je tato variabilita implikovaná modelem spočítána a zanesena do grafů v obrázku III.7. Pruh znázorňuje modelované hodnoty prostředních 68 % (+- 1 SD v normálním rozložení) členů latentní třídy. Skutečně pozorované hodnoty členů latentní třídy by byly leště více rozptýlené, a to o chybu měření, která je v případě tohoto modelu 1,6 (tedy podobná jako SD počátečních stavů). Vedle výše popsané majoritní třídy extrahoval model menší třídu o relativní četnosti 10 %, která reprezentuje adolescenty s rychlejším růstem užívání návykových látek. Na počátku ve 13 letech je si tato třída blízka s majoritou – počáteční hodnoty se liší jen o 2,2, přičemž směrodatná odchylka počáteční hodnoty v této skupině je 2,4. I na grafu je vidět, že se počáteční hodnoty dvou tříd překrývají. Díky podstatně rychlejšímu růstu již mezi 13. a 15. rokem (daným posunutými časovými bázemi spíše než směrnicí) se od majoritní třídy rychle vzdálí a v 15. letech by členové těchto dvou tříd měli být

jasně odlišitelní. Ve třídě s rychlým růstem totiž většina členů udává v 15. letech vyšší míru užívání, než je průměr v 19. Dále již však růst nepokračuje a vysoká míra užívání se příliš nemění; dokonce mezi 17. a 19. rokem model predikuje mírný pokles. V 19 letech již se také hodnoty členů třídy s rychlým růstem překrývají s hodnotami majoritní třídy.

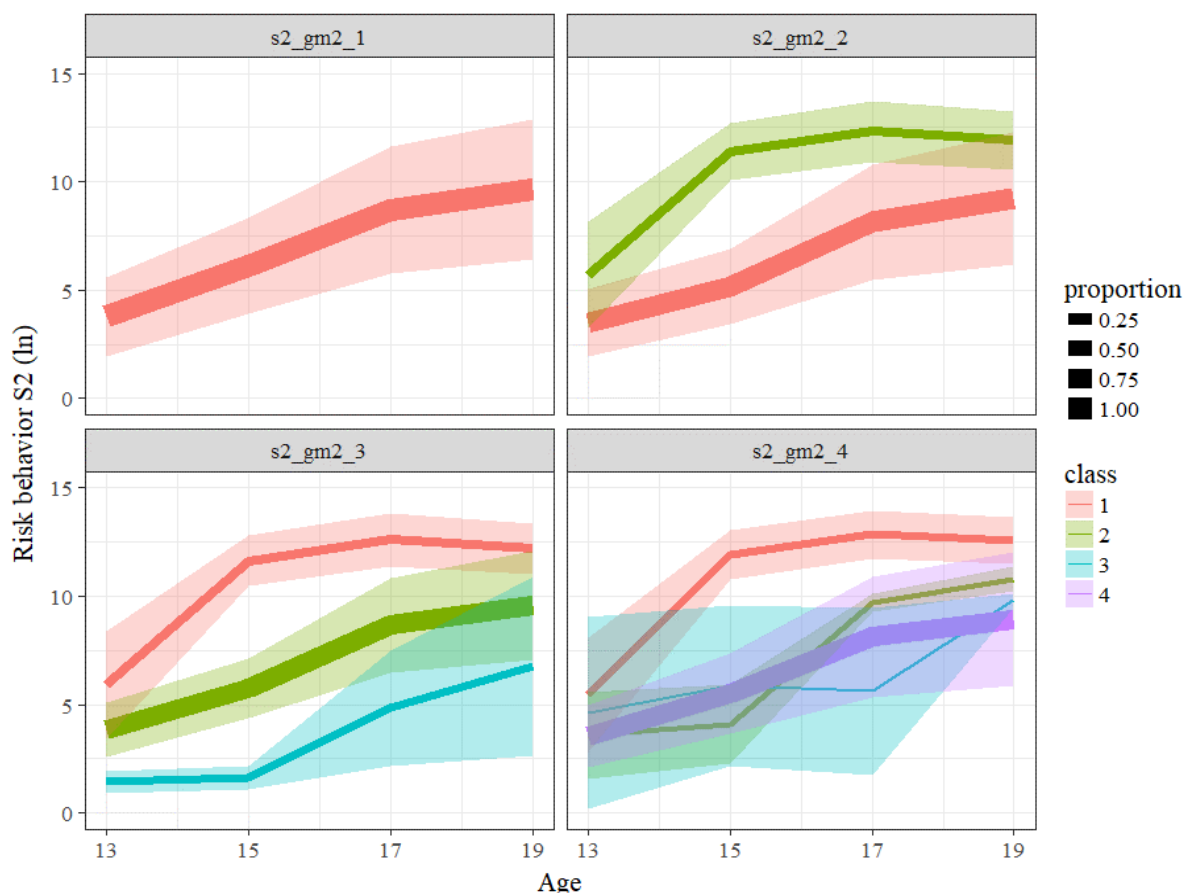
Tabulka III.10. GMM modely užívání návykových látek s časovými bázemi různými napříč třídami.

Parametr		C	1	SE	2	SE	3	SE	4	SE
Relativní četnosti tříd		1	1		0,90		0,12		0,09	
		2			0,10		0,80		0,04	
		3					0,08		0,02	
		4							0,85	
Průměry růstových parametrů	I	1	3,77	0,12	3,50	0,13	5,92	0,50	5,46	0,48
	S	1	0,98	0,04	0,96	0,04	1,05	0,13	1,19	0,19
	I	2			5,71	0,48	3,82	0,17	3,58	0,97
	S	2			1,03	0,12	0,96	0,04	1,21	0,27
	I	3					1,43	0,61	4,60	0,98
	S	3					0,89	0,17	0,87	0,11
	I	4							3,54	0,18
	S	4							0,90	0,06
Rozptyly Růstových parametrů	I	1	3,36	0,55	2,39	0,34	6,14	1,92	6,81	2,25
	S	1	0,14	0,03	0,19	0,04	0,17	0,07	0,16*	0,04
	I	2			6,00	1,84	1,56	0,31	3,99	2,02
	S	2			0,16	0,06	0,16	0,04	0,16*	0,04
	I	3					0,27	0,65	19,52	2,73
	S	3					0,45	0,13	0,16*	0,04
	I	4							2,04	0,36
	S	4							0,16*	0,04
Kovariance růstových parametrů	I-S	1	0,17	0,10	0,02	0,09	-0,90	0,31	-0,94	0,25
	I-S	2			-0,82	0,32	-0,08	0,12	-0,78	0,35
	I-S	3					0,05	0,19	-2,09**	0,30
	I-S	4							0,15	0,10
Reziduální rozptyly Časové báze (Věk-13)	13-19		3,19	0,23	2,42	0,19	2,28	0,19	2,14	0,20
	15 1		2,41	0,15	1,77	0,16	5,48	0,60	5,45	0,58
	17 1		5,06	0,15	4,86	0,16	6,41	0,63	6,23	0,58
	15 2				5,52	0,56	2,03	0,18	0,43	0,49
	17 2				6,41	0,58	5,05	0,21	5,06	0,54
	15 3						0,20	0,35	1,47	0,24
	17 3						3,82	0,68	1,18	0,25
	15 4								2,21	0,22
17 4								5,11	0,23	

Poznámka. I - intercept, počáteční stav, S - směrnice, tempo růstu za rok, C - latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0, červeně problematické parametry. Čas je kódovaný jako Věk - 13, časové báze pro věk 15 a 17 jsou volně odhadované. *Rozptyly směrnic jsou zafixovány na stejnou hodnotu napříč třídami pro dosažení konvergence. **Nepřípustný odhad; hodnota kovariance převyšuje součin směrodatných odchylek.

Model se třemi třídami vyděluje z majority navíc další malou třídu (8 %), a to adolescenty, kteří udávají na počátku pouze minimální užívání návykových látek. Díky tomu mírně klesla variabilita růstových křivek v majoritní třídě (SD počátečních hodnot = 1,2 a SD směrnice = 0,40). I v této třídě „počátečních neuživatelů“ je trend růstový, ale až od 15. roku. Tato třída má jen minimální rozptyl počátečních hodnot užívání (SD = 0,52), ale poměrně vysoký rozptyl směrnice (SD = 0,67). V důsledku toho spolu s růstem průměrné míry užívání roste také variabilita hodnot užívání. Ta je v 19 letech velmi vysoká a hodnoty členů třídy v 19 letech se podobají hodnotám členů ostatních tříd, byť zůstávají v průměru nižší. Z perspektivy modelu se 3 třídami tak existují vedle majority s postupně rostoucí mírou užívání také subpopulace s brzkým nárůstem užívání, jehož intenzita jen zřídka klesne zpět do průměru a subpopulace s pozdním nástupem užívání až po 15. roce, u níž však může vývoj mezi 17. a 19. rokem pokračovat jakkoli. Tato nejistota ohledně pozdního vývoje má na svědomí mírný pokles entropie v modelu se 3 třídami.

Obrázek III.7. Růstové trajektorie jednotlivých latentních tříd v GMM modelech s 1 až 4 latentními třídami s variabilními časovými bázemi.



Poznámka. Plouščka trajektorie vyjadřuje relativní četnost dané latentní třídy. Průsvitný pruh kolem křivky znázorňuje variabilitu růstových křivek uvnitř dané třídy – jednu směrodatnou odchylku nad a jednu pod průměrnou trajektorií. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

Model se čtyřmi třídami poměrně jasný obrázek načrtnutý modelem se 3 třídami rozměluje. Zachována zůstává třída ranných intenzivních uživatelů. Majoritní třída roste na 85 % a z ní se vydělují dvě velmi malé třídy, které se od ní neodlišují počáteční mírou užívání, ale spíše tím, ve kterém období nastává největší nárůst míry užívání – v jedné je to mezi 15. a 17. rokem a v druhé mezi 17. a 19. rokem. Obě jsou však velmi malé (2 a 4 %) a při započtení

variability růstových křivek uvnitř tříd se s majoritní třídou velmi překrývají. Toto překrývání se projevuje také nezanedbatelným poklesem entropie.

Za optimální zde považuji model se třemi třídami, byť právě tento model nedoporučuje žádné ze statistických kritérií. Zachovává však většinu prvků modelu se dvěma třídami, která je optimální podle BIC, BLRT i entropie. Rozšíření o málo četnou třídu „počátečních neuživatelů“ je sice statisticky nepodpořeno, nejspíše kvůli malé velikosti vzorku, ale teoreticky bychom jej čekali. Navíc pokles entropie není až tak velký.

Krok 4 – Volba optimálního modelu

Po představení řady modelů, které různým způsobem popisují heterogenitu vývojových trajektorií, je na místě zvážit, který z nich ji popisuje nejlépe. Je to LGC model, který veškerou heterogenitu popisuje jako normální rozložení růstových křivek (resp. jejich růstových parametrů), nebo LCGA model, který heterogenitu popisuje mnoha vnitřně homogenními třídami, anebo GMM model, který heterogenitu popisuje menším množstvím tříd s nějakou vnitřní mírou heterogenity růstových křivek? Výchozí model latentních růstových křivek odpovídá datům velmi dobře a žádný z modelů latentních růstových tříd (LCGA) jej v tomto ohledu nepřekonává, vezmeme-li v potaz komplexitu modelu. Je to jednak díky tomu, že je patrně do značné míry naplněn předpoklad normality rozložení růstových parametrů. Neobvyklých růstových trajektorií je poměrně málo, aby jejich osamostatnění vedlo k zásadnímu vylepšení modelu. Navíc, teorie nepostuluje existenci několika jasně stanovených trajektorií vývoje užívání návykových látek, vyjma snad jedné, a tou je ranné započetí užívání, které si po zbytek adolescence (i později) udržuje vysokou míru užívání. Přesto je vylepšený model s 5 latentními třídami použitelný, protože s rozumnou mírou entropie umožňuje klasifikaci a další analýzy vyplývající z této klasifikace. Lze například samostatně pracovat jen s vybranými třídami, pokud bychom nebyli ochotni uvěřit v autentičnost malé třídy adolescentů stabilně ve vysoké míře užívajících návykové látky od 13 do 19 let.

Modely růstových směsí (GMM) přeci jenom dokáží shodu modelu s daty oproti LGC modelu o něco málo vylepšit. Záleží však na dobré specifikaci. Zatímco model, v němž byl základní tvar všech vývojových trajektorií stejný, nenabídl žádné použitelné řešení, model s variabilním tvarem růstových trajektorií napříč třídami nabídl zajímavé a teoreticky smysluplné řešení. I jeho konzervativnější varianta se dvěma latentními třídami je dobře použitelná. Pro následné analýzy, v nichž nás zajímají jak koreláty členství ve třídě, tak koreláty růstových parametrů uvnitř tříd je tak GMM model se dvěma nebo třemi třídami velmi dobrou výchozí pozicí.

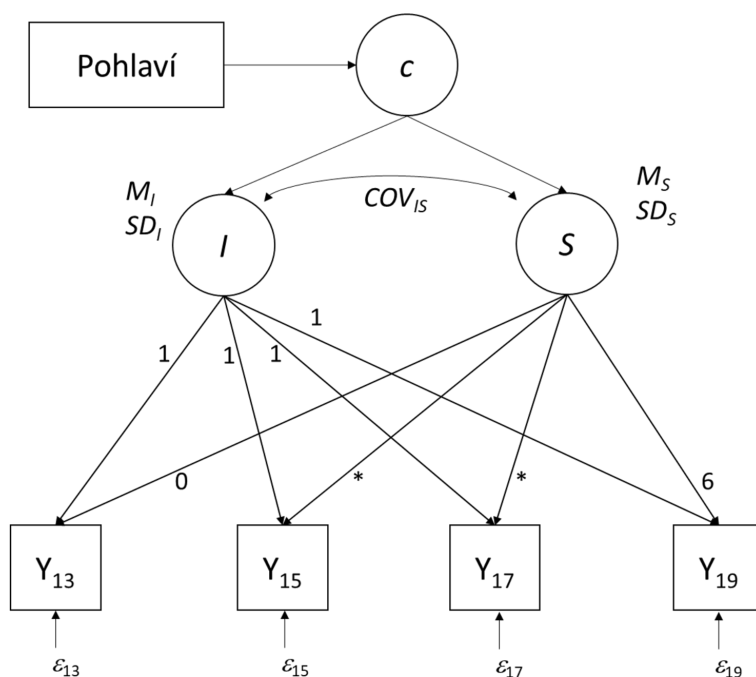
Věcně modely identifikovaly dva nejčtenější typy vývoje – postupný mírně nelineární nárůst užívání mezi 13. a 19. rokem a intenzivní užívání s brzkým nástupem mezi 13. a 15. rokem. Tyto dva druhy vývoje se vynořovaly napříč většinou modelů a je na místě uvažovat, že jako subpopulace reálně existují. Dle teorie by měla existovat také subpopulace vyhýbající se návykovým látkám, ale ta byla v našich datech tak málo reprezentována, že její statistická podpora není úplně přesvědčivá.

Krok 5 - Zařazení kovariátů do GMM modelů

Pro demonstraci zařazení kovariátů do GMM modelu jsem vybral GMM model se dvěma třídami s odlišným tvarem trajektorie popsany v předchozí části. Vhodným kovariátem pro tuto demonstraci se zdá proměnná pohlaví, protože jde o dichotomii kódovanou 0 – chlapci a 1- dívky. Díky tomu lze s proměnnou pracovat jako s kategoričkou i spojitou a výsledky mezi různými analytickými přístupy snadno srovnat. Pohlaví může v kontextu GMM modelu užívání návykových látek mít roli prediktoru členství ve třídě a roli prediktoru latentních růstových koeficientů (počátečního stavu a tempa růstu) uvnitř třídy. Lze jej technicky v modelu použít také jako závislou proměnnou, ovšem věcně je to nesmysl.

Jak je uvedeno v teoretické části práce, můžeme pohlaví jako prediktor přímo specifikovat jako součást modelu (spojitou, či kategoričkou), nebo můžeme ve snaze vyhnout se nebezpečí změny modelu měření latentních tříd, použít Vermuntův tříkrokový postup, který odděluje odhad nepodmíněného GMM modelu a modelu s prediktorem. Tento tříkrokový postup lze realizovat ručně, nebo využít automatické funkce Mplus. I když se to nedoporučuje, lze využít také přístup, který opakovaně náhodně přiřazuje jedince do tříd podle klasifikačních pravděpodobností a na takto opakovaně generovaných datech zjišťuje efekt prediktoru (*pseudo-class* přístup).

Obrázek III.8. GMM model užívání návykových látek s prediktorem členství ve třídě



Jednokrokový postup

Proměnnou pohlaví učiníme spojitým prediktorem latentní třídy pomocí logistické regrese. V Mplus prostým zařazením `C ON sex;` v `%OVERALL%` části specifikace modelu. Proměnná pohlaví je bez dalšího upřesnění považována za spojitou. Mplus v této roli nepřipouští kategoričké proměnné – ty je případně potřeba ručně transformovat na dummy proměnné.

Výstup Mplus ohledně jednotlivých latentních tříd zůstal strukturně nezměněn. K posunu parametrů tříd nedošlo (viz tabulka II.7 výše), pravděpodobnosti členství byly ovlivněny minimálně a entropie se minimálně posunula o jednu tisícinu na 0,764. Efekt pohlaví na členství ve třídě je uveden v sekci *Categorical Latent Variables* v logitové metrice. Uvádí-li tedy výstup hodnotu parametru -0,041, znamená to, že jednotkový rozdíl hodnoty prediktoru odpovídá poklesu logaritmu poměru šancí (logitu) členství v 1. třídě (majorita) oproti 2. třídě (rychlý růst) o 0,041. To se samo obtížně interpretuje, a proto je potřeba tento výsledek převést na poměr šancí či pravděpodobnosti. Jde o běžnou logistickou regresi, takže platí, že poměr šancí je dán exponenciální funkcí: $e^{-0,041} = 0,960$. Vzhledem k tomu, že pohlaví je kódováno 1 pro dívky a 0 pro chlapce a referenční třídou je druhá (poslední třída), dozvídáme se, že dívky mají 0,96krát ($e^{-0,041}$) vyšší šanci být v první třídě než ve druhé třídě. Zatímco pro chlapce je $P(C=1 | \text{sex}=0) = e^{1,815} / (1 + e^{1,815}) = 0,86$, pro dívky je $P(C=1 | \text{sex}=1) = e^{1,815-0,041} / (1 + e^{1,815-0,041}) = 0,85$, což je zanedbatelný a nesignifikantní rozdíl ($p=0,90$). Převod z hodnoty parametru na poměr šancí je již součástí výstupu. Součástí výstupu je také přepočítání pro situaci, kdyby referenční třídou byla první třída. Například v kontextu aktuálního GMM modelu se zdá být smysluplnější klást otázku tak, že se ptáme, zda se některé pohlaví nepojí s větší pravděpodobností členství v 2. latentní třídě, tedy třídě s časným a rychlým nástupem užívání návykových látek. Efekt pohlaví na členství ve třídě je však minimální a nesignifikantní, což je z pohledu teorie poněkud překvapivé.

```

Categorical Latent Variables
C#1      ON
      SEX          -0.041      0.320      -0.128      0.898
Intercepts
C#1          1.815      0.324      5.593      0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS
Categorical Latent Variables
C#1      ON
      SEX          0.960

ALTERNATIVE PARAMETERIZATIONS FOR THE CATEGORICAL LATENT VARIABLE REGRESSION
Parameterization using Reference Class 1
C#2      ON
      SEX          0.041      0.320      0.128      0.898
Intercepts
C#2          -1.815      0.324      -5.593      0.000

```

Tříkrokový postup

Při použití automatického tříkrokového postupu pro testování efektu prediktorů na členství v latentní třídě nestanovujeme efekt pohlaví na členství ve třídě jako součást modelu. Jeho specifikace zůstává z nepodmíněného GMM modelu nezměněna. Slouží k tomu příkaz `AUXILIARY = SEX (R3STEP) ;`, který je uveden v sekci `VARIABLE :`. Jeho uvedení instruuje Mplus, aby realizoval všechny tři kroky a v závěrečném specifikoval logistickou regresi latentní třídy C na pohlaví. Výstup Mplus obsahuje popis nepodmíněného GMM modelu a samostatně logistickou regresi třídy na pohlaví ve stejném formátu, jako v jednokrokovém řešení.

TESTS OF CATEGORICAL LATENT VARIABLE MULTINOMIAL LOGISTIC REGRESSIONS USING THE 3-STEP PROCEDURE

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
C#1	ON				
	SEX	0.046	0.350	0.131	0.896
	Intercepts				
	C#1	1.768	0.247	7.154	0.000
Parameterization using Reference Class 1					
C#2	ON				
	SEX	-0.046	0.350	-0.131	0.896
	Intercepts				
	C#2	-1.768	0.247	-7.154	0.000

Výsledek se od jednokrokového liší v řádu setin se stejným závěrem – efekt pohlaví na členství ve třídě je zanedbatelný a statisticky nesignifikantní. Na rozdíl od jednokrokového postupu není součástí výstupu poměr šance na členství v první třídě mezi chlapci a dívkami, takže si jej v případě potřeby musíme spočítat: $e^{0,046} = 1,05$. Ostatní části modelu jsou identické s modelem bez kovariátů. Opět je uvedena i alternativní parametrizace užívající první třídu jako referenční.

Manuální tříkrokový postup vede v tomto případě ke zcela stejným odhadům, jako automatický tříkrokový postup R3STEP. V prvním kroku je potřeba spustit odhad nepodmíněného modelu, kdy zařazení příkazu

```
SAVEDATA: FILE = "predikovane_clenstvi.dat";
SAVE = CPROB;
```

uloží do nového datového souboru pravděpodobnosti členství v jednotlivých třídách a nejpravděpodobnější třídu pro všechny respondenty. Aby se do dat uložily i hodnoty identifikační proměnné, je potřeba do sekce VARIABLE: uvést, která proměnná to je, např. IDVARIABLE = id;. Z výstupu tohoto modelu také vyčteme souhrnné logity členství ve jednotlivých latentních třídách. Z následující tabulky:

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)by Latent Class (Row)

	1	2
1	4.005	0.000
2	-0.506	0.000

použijeme logity pro specifikaci dummy modelu latentních tříd, která už nebude závislá na indikátorech – měřeních užívání návykových látek. Proměnnou, která nese informaci o nejpravděpodobnější třídě je ještě potřeba deklarovat jako nominální pomocí NOMINAL= class; v sekci VARIABLE:.

```
VARIABLE: NAMES = id n sex cprob1 cprob2 class;
          NOMINAL= class; !Stanovíme nejpravděpodobnější třídu jako nominální.
          MISSING = ALL (999);
          USEVARIABLES = class;
          IDVARIABLE = id;
          CLASSES = C(2);

MODEL:
%OVERALL%
%c#1%
[class#1@4.005];
%c#2%
[class#1@-0.506];
```

Po spuštění dummy modelu a ujištění se, že se model měření tříd oproti původnímu GMM modelu nezměnil, můžeme do modelu přidat regresi třídy na pohlaví. Stojí za zmínku, že v našem případě, ač zůstaly relativní četnosti tříd a klasifikační pravděpodobnosti zachovány, klesla hodnota entropie v dummy modelu na 0,64 (oproti 0,76 u výchozího modelu).

```
MODEL:
  %OVERALL%
  C ON sex;

  %c#1%
  [class#1@4.005];

  %c#2%
  [class#1@-0.506];
```

Výstup pak je strukturován stejně jako v předchozích způsobech provedení této analýzy a v tomto případě obsahuje také zcela stejné odhady efektu pohlaví na latentní třídu, jako automatický tříkrokový postup R3STEP.

Categorical Latent Variables

C#1	ON				
SEX		0.046	0.350	0.131	0.896
Intercepts					
C#1		1.768	0.247	7.154	0.000

LOGISTIC REGRESSION ODDS RATIO RESULTS

Categorical Latent Variables

C#1	ON	
SEX		1.047

ALTERNATIVE PARAMETERIZATIONS FOR THE CATEGORICAL LATENT VARIABLE REGRESSION

Parameterization using Reference Class 1

C#2	ON				
SEX		-0.046	0.350	-0.131	0.896
Intercepts					
C#2		-1.768	0.247	-7.154	0.000

V případě GMM modelu se dvěma třídami můžeme tedy konstatovat, že pohlaví nesouvisí s pravděpodobnostním členstvím v latentních třídách. Různé přístupy k posouzení prediktivního efektu pohlaví, jednokrokové či vícekové, v tomto případě poskytují velmi podobné výsledky.

Latentní třídy vývoje užívání návykových látek

Uvedená paleta modelů vývoje užívání návykových látek během adolescence v projektu ELSPAC nabízí řadu podnětů k uvažování; nelze však říci, že by z nich plynul jasný závěr o existenci subpopulací s odlišným vývojem. Výchozí model latentních růstových křivek je poměrně dobrým popisem heterogenity vývojových křivek. Modely s růstově vnitřně homogenními latentními třídami tak data nepopisují lépe než on. Třídy, které LCGA modely produkují, jsou spíše diskretizací spojitého rozložení, ať již je modelu umožněno měnit tvar růstových křivek napříč třídami či ne. Na druhou stranu, modely s různými tvary růstové křivky napříč třídami již naznačovaly to, co se pak vynořilo v růstových mixture modelech,

tedy smysluplnou klasifikaci růstových křivek podle věku nástupu užívání. Časný nástup užívání byl v minulosti mnohokrát spojen s vyšší intenzitou užívání (např. Gruber, DiClemente, Anderson, & Lodico, 1996), a proto jsou s různou dobou nástupu užívání spojeny jiné tvary vývojových trajektorií.

Růstový mixture model poskytoval smysluplná řešení je, pokud mu bylo dovoleno uvolnit tvar růstových křivek napříč třídami. Pak identifikoval tři třídy, které jsou v souladu s teorií: majoritní třídu, v níž postupně mezi 13 až 19 lety roste míra užívání, přičemž nejvyšší nárůst je mezi 15 a 17 lety, minoritu s brzkým nástupem užívání ve 13 letech a celkově vysokou mírou užívání a minoritu, která až do 17 let neudává téměř žádné užívání. Zůstává otázkou, jaký je v poslední třídě vývoj mezi 17 a 19 lety, protože tam již měl model k dispozici velmi málo dat, a tak je v této periodě rozptýl užívání odhadnutý poměrně vysoko. Normativnost užívání návykových látek v naší kultuře (Širůček, Širůčková, 2008) je však z celkového modelu poměrně zjevná, protože všechny růstové trajektorie směřují vzhůru a postupně konvergují.

Zjištěním, které zvyšuje skepsi ohledně platnosti zredukované verze této klasifikace na latentní třídy, je absence genderových rozdílů mezi třídami. I když se rozdíly mezi chlapci a děvčaty v užívání alkoholu v adolescenci postupně zmenšují, stále ještě existují (Csémy, Hamanová, 2009) a měly by se tedy projevit i v našich datech.

IV. ANALÝZA HETEROGENITY VÝVOJOVÝCH KŘIVEK PROŽÍVANÉ AUTONOMIE V MLADÉ DOSPĚLOSTI

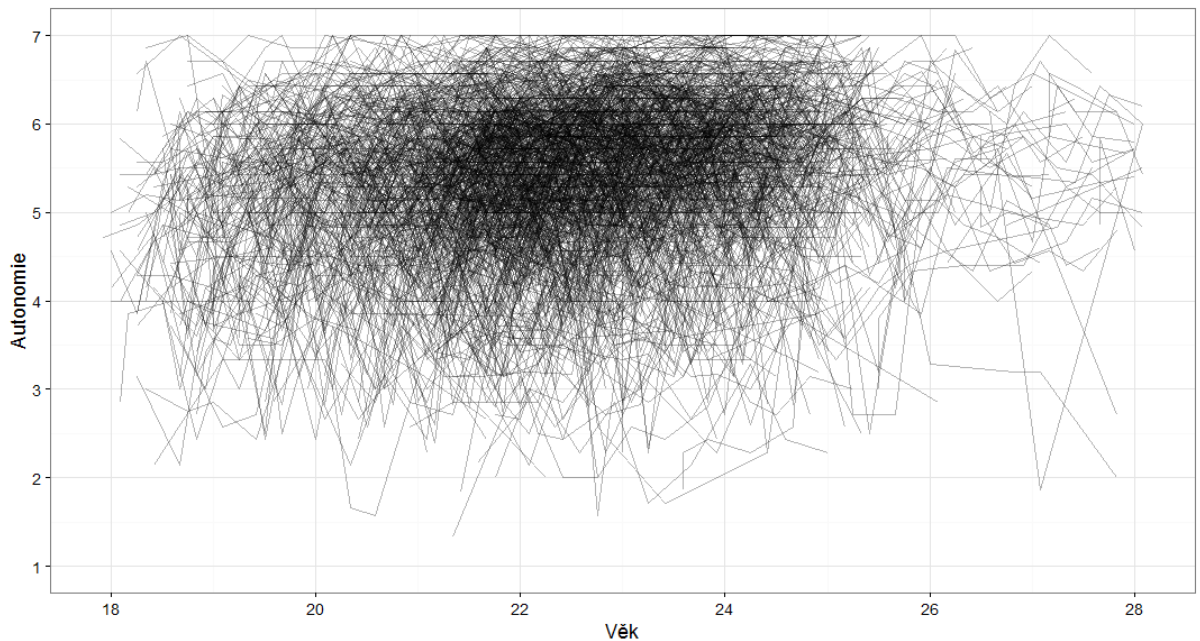
V rámci projektu *Cesty do dospělosti* (Ježek, Macek, Bouša, 2016) jsme průběžně sledovali subjektivně prožívanou autonomii (cf Ježek, 2014). V každé ze 13 vln měření během čtyřletého období jsme účastníkům (N = 1751) studie administrovali sedmipoložkovou škálu prožívané autonomie, která je subškálou české adaptace nástroje *Basic Psychological Needs Satisfaction Scale* (Deci & Ryan, 2000). Každá položka (např. Můžu se sám/sama rozhodovat, jak chci žít svůj život.) nabízí výběr ze sedmi stupňů souhlasu kódovaných do 1 do 7. Jde o Likertovskou škálu, kde je však celkový skóre počítán jako průměr položkových odpovědí, nikoli součet, pro usnadnění interpretace. Vyšší hodnoty znamenají vyšší úroveň prožívané autonomie. Na počátku studie bylo účastníkům projektu mezi 18 a 24 lety, a tak data pokrývají období od 18 do 28 let. Škálu lze považovat za jednodimenzionální (viz longitudinální CFA níže v sekci o růstových modelech druhého řádu) s vnitřní konzistencí pohybující se v různých věcích a při různých administracích kolem hodnoty 0,7.

Otázky, kterou můžeme klást takovým datům, jsou primárně, jakou podobu mají individuální vývojové trajektorie autonomie, nakolik se mezi lidmi liší, a zda v nich existují nějaké pravidelnosti. Sekundárně se můžeme ptát po prediktorech či korelátech průběhu vývoje autonomie. Můžeme se například ptát, zda se liší průměrné trajektorie různých skupin, např. mužů a žen, nebo vývojově přiléhavěji těch, kdo nastoupili na VŠ a těch, kdo ne. Mohou nás při tom zajímat vlivy korelátů, jako je například aktuální nálada či jednorázových vlivů, jako je například významná pozitivní či negativní životní událost.

V první řadě je ale na místě se seznámit s individuálními vývojovými trajektoriemi. Ty, pro celý vzorek projektu *Cesty do dospělosti*, zobrazuje obrázek IV.1. Odpovědi jednotlivých účastníků jsou spojeny (interpolovány) pro toto zobrazení úsečkami, takže body, v nichž se čára láme jsou hodnoty autonomie naměřené v daném věku. Toto zobrazení zahrnuje všechny účastníky, kteří odpovídali na škálu autonomie alespoň čtyřikrát (N = 1012). Protože jednotliví účastníci měli určitou volnost v tom, kdy přesně který dotazník vyplní, a samozřejmě měli volnost v tom, zda vůbec dotazník, popř. konkrétní škálu, otázky zodpoví, je v datech zaznamenáno datum vyplnění, a tedy i věk účastníka v době odpovídání. Proto je osa X spojitá a různé křivky se lámou v různých věcích. Taková časová volnost při sběru dat znamená řadu praktických výhod z hlediska organizace výzkumu, komfortu, a tedy i motivace účastníků. Věřím, že by bylo možné argumentovat, že tak lze získat více a kvalitnějších dat. Z hlediska analýzy dat je však výzvou, jak se ukáže dále.

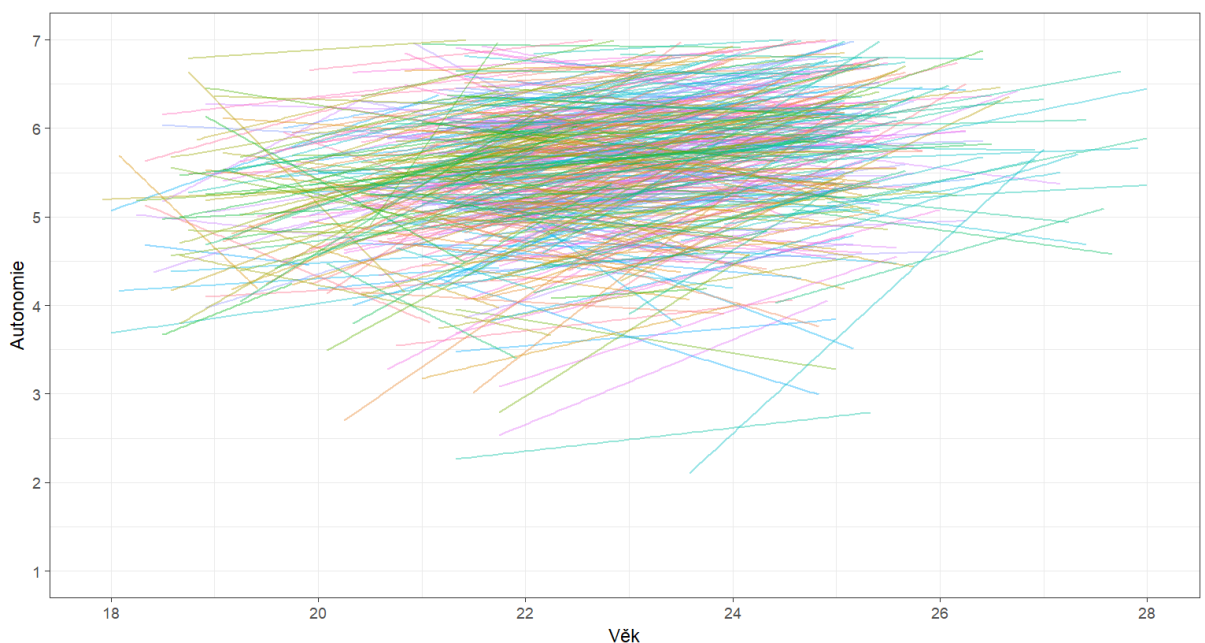
Z pohledu na obrázek IV.1 je zřejmé, že vývojové trajektorie prožívané autonomie jsou velmi různorodé. Individuální trajektorie se často prudce lámou a svědčí o tom, že v konkrétních chvílích se prožívaná autonomie může velmi lišit od předchozích měření. To může být způsobeno skutečnými změnami prožívané autonomie, ale také náhodnou chybou měření, která může mít mnoho podob a příčin. Z tohoto zobrazení není možné přímo usuzovat na to, nakolik je smysluplné předpokládat, že tyto divoce lomené čáry mají obecně spíše setrvalou úroveň, či zda autonomie v rámci toho kolísání mírně roste nebo klesá, popř. zda se tempo změny během času nějak mění (např. zrychluje/zpomaluje).

Obrázek IV.1. Individuální vývojové trajektorie autonomie účastníků s účastí v alespoň 4 vlnách (N = 1012)



Kdybychom například předpokládali, že individuální růst je lineární, mohli bychom proložit individuálními hodnotami každého účastníka výzkumu přímkou (pomocí běžné lineární regrese s kritériem nejmenších čtverců). Výsledkem by bylo zobrazení na obrázku IV.2. Úsečky zde zobrazené začínají i končí ve věcích, kdy se účastník studie skutečně účastnil. Nejsou tedy extrapolované, což nám umožňuje si uvědomit, že díky tomu, že trvání studie bylo kratší než zobrazených 10 let (18–28), mají v mnoha případech úsečky tak prudký sklon, že by to při extrapolaci na celé desetileté rozpětí znamenalo predikované hodnoty mimo rozsah měřicího nástroje. To může být jak projevem náhodné chyby měření ve spojení s malým počtem měření (zde zobrazení respondenti se 4 a více měřeními), i tím, že vývoj autonomie očistěný od náhodných, situačních vlivů nemusí být v celém věkovém rozpětí let lineární.

Obrázek IV.2. Odhadnuté individuální lineární trajektorie



Umožňuje to také si všimnout, že většina úseček má stoupající trend, zdaleka však ne všechny. Co však neumožňuje posoudit, je to, do jaké míry lze považovat různost individuálních trendů za projev náhody/chyby měření či systematických inter-individuálních rozdílů.

Model růstových křivek založený na víceúrovňové regresi umožňuje přímo pracovat proměnnou „věk v době měření“⁶, která nabývá pro různé účastníky výzkumu různých hodnot (tj. různí účastníci byli měřeni v různých věcích), přičemž různí účastníci mohou mít různý počet měření. Data z projektu Cesty do dospělosti tak není potřeba nijak upravovat a pokud jsou v dlouhém formátu, lze růstový model specifikovat například procedurou MIXED v SPSS či jako multilevel lineární model v Mplus.

Jak je patrné i z grafů individuálních vývojových křivek, rozložení proměnné autonomie je zjevně zleva zešíkmené. Za tímto zešíkmením stojí jednak strop krátkého, pouze sedmipoložkového měřicího nástroje a pak také fakt, že ve třetí dekádě mají lidé objektivně vysokou míru autonomie, volnosti, což subjektivní prožívání autonomie do značné míry reflektuje. Různé metody, které v této práci využívám, jsou na takovéto odchylky od normality různě citlivé (zejména mixture modely), a proto je vhodné non-normalitu limitovat. I když existují sofistikovanější způsoby vypořádání se odchylky od normality modelované proměnné (zejména generalizované lineární modely), so ohledem na srovnatelnost napříč paletou růstových modelů, jsem zde použil jednoduchou kvadratickou transformaci proměnné autonomie. Ta koriguje zešíkmení, nikoli však efekt stropu, a zachovává interpretovatelnost skóre.

Tabulka IV.1 prezentuje souhrn parametrů tří nepodmíněných modelů růstových křivek – konstantního (žádná změna), lineárního a kvadratického. Modely byly odhadnuty pomocí procedury MIXED v IBM SPSS v. 24 (IBM, 2016). Pro usnadnění interpretace byl věk v těchto modelech centrován na 18. rok, tj. od věku je odečtena hodnota 18. První model – model stability – je specifikován tak, že jednotlivci se mohou lišit svou hodnotou autonomie, ale ta se s věkem nemění. V tomto modelu jsou tedy vývojovými trajektoriemi vodorovné přímky (individuální průměrné autonomie), které jsou normálně rozloženy s průměrnou hodnotou kvadratické autonomie 29,2 a směrodatnou odchylkou 7,2. Reziduální rozptyl má hodnotu 35,4, což znamená, že naměřené hodnoty jednotlivce kolísají kolem jeho individuálního průměru se směrodatnou odchylkou 5,9. Ta je tedy jen o málo nižší, než je inter-individuální variabilita. Poměr variability mezi jednotlivci a intra-individuální variability můžeme vyjádřit také vnitrotřídním korelačním koeficientem, který zde nabývá hodnotu $51,35 / (51,35 + 35,37) = 0,59$. Můžeme tedy říci, že necelých 60% rozptylu naměřených hodnot lze vysvětlit normálně rozloženými interindividuálními rozdíly v autonomii. To je poměrně mnoho, což podporuje hypotézu, prožívání autonomie je poměrně stabilní individuální charakteristika, i když se to tak pohledem na graf individuálních trajektorií nemusí zdát.

⁶ Mplus to v LGC modelu umí také, a to pomocí TSCORES. Pořád využívá široká data, ale k jednotlivým měřením přidává informaci o věku v době měření. Bohužel pak ve výstupu chybí velká část ukazatelů shody modelu s daty.

Tabulka IV.1. Shrnutí parametrů modelů růstových křivek (GCM) autonomie

	Stabilní autonomie, náhodný intercept	Lineární růst s náhod- nými parametry	Kvadratický růst s ná- hodnými parametry
Růstové parametry			
Počáteční hodnota	29,21	24,12	23,01
(Věk-18)		1,16	1,75
(Věk-18) ²			-0,07
Rozptyly			
Rezidua	35,37	29,20	28,42
Počáteční hodnota (I)	51,35	93,92	78,93
(Věk-18)		2,54	12,98
(Věk-18) ²			0,11
COV (I; (Věk-18))		-10,45	-18,07
COV (I; (Věk-18) ²)			-1,07*
COV ((Věk-18); (Věk-18) ²)			0,69
Odvozené parametry			
ICC	0,59	0,66	0,67
r (I; (Věk-18))		-0,68	-0,56
r (I; (Věk-18) ²)			0,23
r ((Věk-18); (Věk-18) ²)			-0,90
Informační kritéria			
-2 Log Likelihood	69764	68774	68655
AIC	69770	68786	68669
BIC	69792	68829	68720
Počet parametrů	3	6	10
Test modelů			
rozdíl -2LL		990,4	118,9
rozdíl df		3	4
p		<0,001	<0,001

Poznámka. (Věk-18) je věk centrovaný na 18. Závislá proměnná autonomie je umocněná na druhou. * p>0,05. Tabulka prezentuje ML odhady parametrů.

Lineární model růstových křivek k předchozímu modelu přidává možnost růstu. Aktuální hodnota kvadratické autonomie tak závisí nejen na nějaké individuální (rysové) hodnotě prožívané autonomie, ale také na věku. Likelihood ratio test ukazuje, že tento růstový model vykazuje lepší shodu s daty ($\chi^2(3) = 990, p < 0,001$). Takový závěr podporuje také vyšší

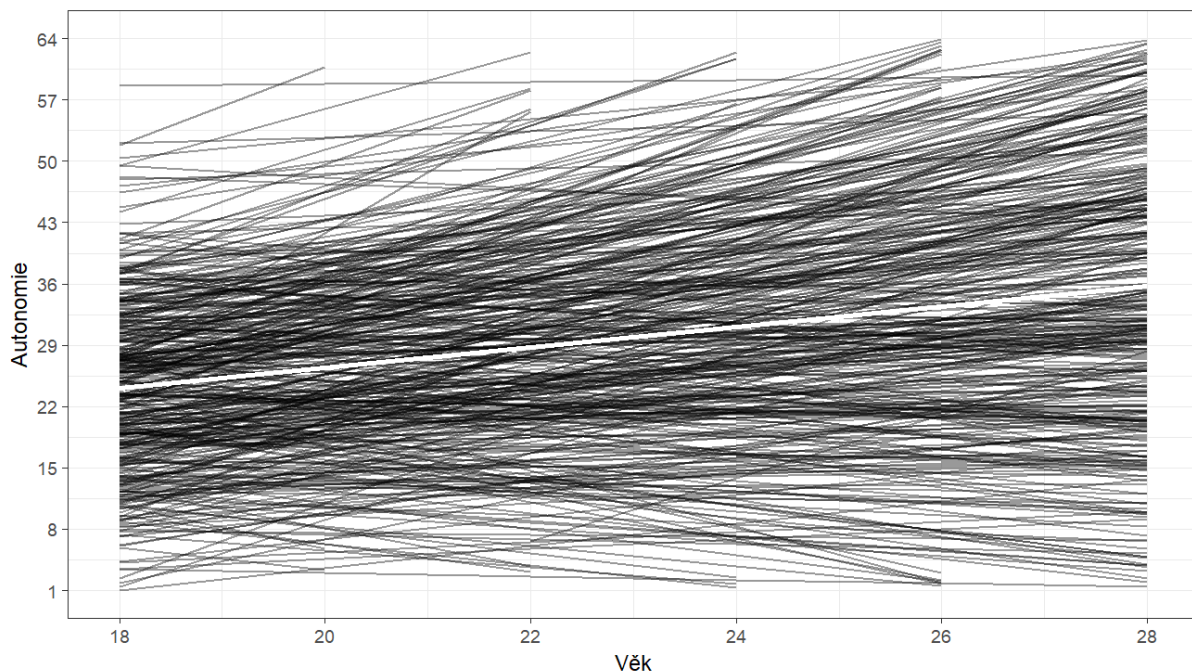
hodnota vnitrotřídní korelace (66%) a nižší hodnota reziduálního rozptylu. Můžeme tedy říci, že model lineární změny autonomie popisuje věrněji to, co vidíme v datech. Jaká ale tato lineární změna je? Průměrná hodnota ročního nárůstu autonomie je 1,16. To není mnoho, zvláště, když si uvědomíme, že pracujeme s kvadratickými hodnotami autonomie. Ale za 10 let to znamená na původní sedmibodové škále autonomie nárůst o 1 bod, tedy šestinu škály. Podstatnější než průměrný nárůst je však to, jak moc se v rychlosti růstu prožívané autonomie jednotlivci mezi sebou liší. Rozptyl parametru růstu je 2,54. Růst má tedy průměr 1,16 ročně se směrodatnou odchylkou 1,59. Předpokládáme-li tedy normální rozložení tohoto růstového parametru, model předpovídá, že asi čtvrtina populace má růstový parametr rovný 0, nebo menší. A zhruba desetina má růstový parametr -1,16 nebo menší – tedy na kvadratické škále autonomie klesá stejným tempem, jakým průměr populace stoupá. Je zřejmé, že nalézáme značné rozdíly mezi lidmi v tom, jak a jakým směrem se jejich autonomie pohybuje. Zůstávají zde samozřejmě i rozdíly mezi jednotlivci v tom, na jaké úrovni autonomie v 18 letech začínají. Modelem implikované růstové přímky mají průměrnou hodnotu ve věku 18 let 24,1 s velkou směrodatnou odchylkou 9,7. Posledním zbývajícím parametrem modelu je kovariance mezi počáteční úrovní autonomie a tempem růstu. Ta je záporná a převedena na korelaci má poměrně vysokou hodnotu -0,68. To znamená, že čím vyšší je počáteční úroveň autonomie, tím nižší je pravděpodobně tempo růstu. Naopak, nižší hodnoty počáteční úrovně autonomie se pojí spíše s vyšším tempem růstu. Takový model tedy popisuje postupné sblížení hodnot prožívané autonomie mezi lidmi.

S ohledem na nepřekročitelnou maximální hodnotu použité měřicí škály i na životní změny, které přichází přibližně kolem 24. roku společně s ukončením vysoké školy se nezdá realistické, že by autonomie napříč celým obdobím od 18 do 28 let rostla nebo klesala lineárně. Kvadratický růst umožňuje postupně zpomalující/zrychlující se tempo změny. Kvadratický model přidává k lineárnímu modelu parametr kvadratického růstu a hned tři další parametry – rozptyl kvadratického růstu, kovarianci kvadratického růstu s počáteční úrovní autonomie a kovarianci kvadratického růstu s lineárním růstem. Díky tomu, že individuální růstové křivky mohou být kvadraticky zakřivené, mohou na naměřené hodnoty lépe pasovat. V tomto případě rozdíl oproti lineárnímu modelu není velký – reziduální rozptyl má hodnotu 28,4 což znamená, že naměřené hodnoty jednotlivce kolísají kolem jeho individuální kvadratické křivky se směrodatnou odchylkou 5,3 (ICC = 0,67). Kvadratické křivky jsou tedy podobně dobrým modelem individuálních vývojových trajektorií jako přímky. Shoda modelu s daty je statisticky významně lepší oproti lineárnímu modelu ($\chi^2(4) = 119, p < 0,001$) a i informační kritéria penalizující za množství volných parametrů (AIC, BIC) mají pro kvadratický model nižší hodnotu a indikují tedy lepší shodu s daty.

Preferenci kvadratického modelu komplikují korelace mezi růstovými parametry. V tabulce IV.1 jsou uvedeny kovariance; převedeno na korelace to znamená -0,56 mezi počáteční úrovní a parametrem lineárního růstu, 0,23 mezi počáteční úrovní a parametrem kvadratického růstu a celých -0,90 mezi parametrem lineárního růstu a parametrem kvadratického růstu. Tyto závislosti značně komplikují možnost představit si rozložení růstových křivek implikovaných modelem. Je zřejmé, že nadprůměrný lineární růst se bude pojit téměř vždy se zápornou hodnotou kvadratického růstu – půjde tedy o zpomalující se růst. Naopak velmi podprůměrný lineární růst se bude pojit s nulovým i mírně pozitivním kvadratickým koeficientem, což reprezentuje lineární až mírně se zrychlující růst. Lepší představu však přinese možnost simulací vybrat z tohoto trojrozměrného normálního rozložení vzorek většího množství křivek a zobrazit je v grafu modelem implikovaných růstových křivek, viz obrázky IV.3 a IV.4 pro lineární a kvadratický model. Tyto grafy

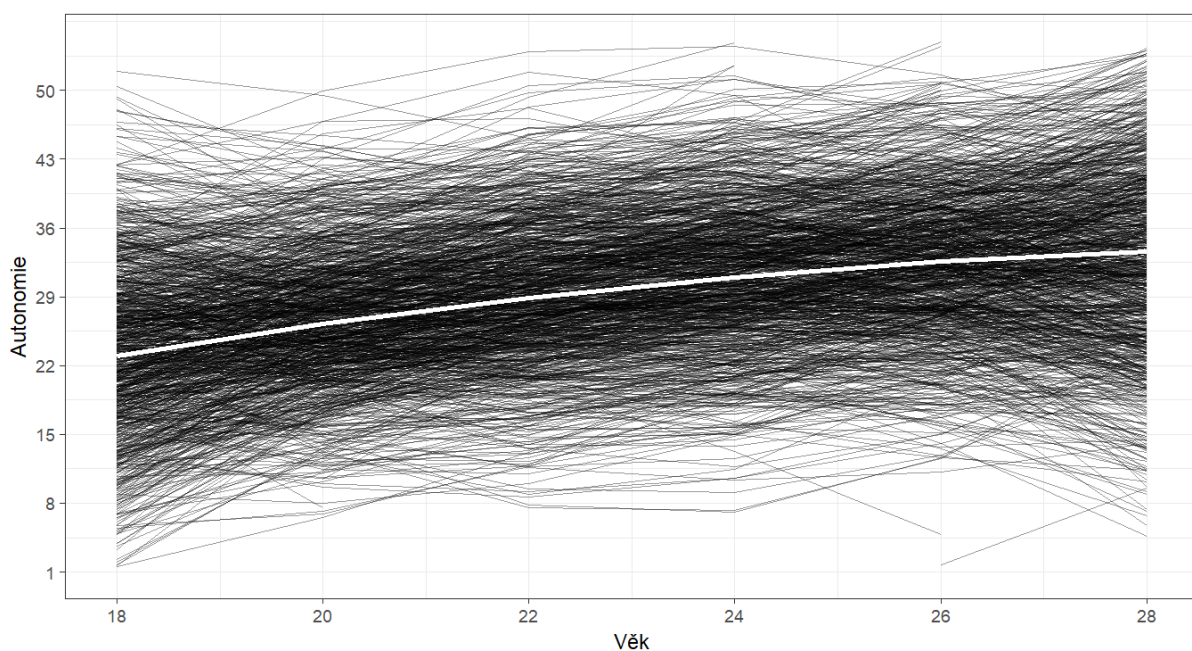
představují zjednodušeně populaci, z níž podle modelů pochází empirické růstové křivky zobrazené na obrázku IV.1 (samozřejmě bez náhodné chyby). Lze si všimnout, že modely nejsou cenzorované, součástí modelu tedy není informace o tom, že závislá proměnná může nabývat pouze hodnot v rozmezí 1 až 49 (autonomie²).

Obrázek IV.3. Rozložení růstových křivek implikovaných lineárním modelem (N=500)



Poznámka. Bílá tučná čára reprezentuje průměrnou vývojovou trajektorii. Strop pro zobrazování přímek je na hodnotě 60, byť maximum škály je 49.

Obrázek IV.4. Rozložení růstových křivek implikované kvadratickým modelem (simulace N=1000)



Poznámka. Bílá tučná čára reprezentuje průměrnou vývojovou trajektorii. Strop pro zobrazování přímek je na hodnotě 60, byť maximum škály je 49.

Na základě výchozích modelů růstových křivek je tedy možné konstatovat, že prožívaná autonomie má silnou stabilní komponentu, protože model bez růstu vysvětluje podstatnou část rozptylu naměřených hodnot. I přes velké individuální rozdíly v úrovni autonomie lze identifikovat růstovou složku. Tu lze modelovat konzervativně jako lineární (konzervativně proto, že teorie neimplikuje komplexnější vývoj) či s ohledem na strop škály jako kvadratický. V obou případech jsou mají růstové komponenty velkou variabilitu a „kompenzující“ vzorec kovariancí. Jak u lineárního, tak u kvadratického modelu jsou mezi jednotlivci velké rozdíly v parametrech růstu, které by mohlo být zajímavé vysvětlit. Mohl bych zde pokračovat rozšířením modelů o možné kovariáty na úrovni jednodlivce, které by mohly vysvětlit rozptyl růstových parametrů, popř. kovariáty či na úrovni jednotlivých měření, které by mohly vysvětlit část reziduálního rozptylu. Tato práce se však věnuje teoretické možnosti, že pozorovaná heterogenita růstu je dána tím, že v populaci existuje několik subpopulací o neznámé velikosti, v rámci kterých je variabilita růstu nižší, nebo zcela nulová. Pojdme se tedy nyní věnovat tomuto způsobu zachycení heterogenity růstu. K tomu však bude nutné výše popsané modely respecifikovat jako modely latentních růstových křivek

Model latentních růstových křivek

PŘÍPRAVA DAT

LGC respecifikace výše uvedených modelů s použitím software Mplus v 7.11 (Muthén & Muthén, 2013) vyžaduje v případě dat o autonomii určitou transformaci dat⁷. Tyto modely na rozdíl od GCM předpokládají, že závislá proměnná byla měřena v pevných (ne nutně stejných) intervalech a že měření v každém jednotlivém čase je zachyceno samostatnou proměnnou. Zatímco u GCM modelů byla závislou proměnnou autonomie a prediktorem věk (který mohl nabývat libovolných hodnot), LGC pracuje s proměnnými zachycujícími naměřenou hodnotu autonomie v určitém věku a proměnná pro věk zde nefiguruje. Jde tedy o dvě podoby dat, mezi kterými je nutné v různých longitudinálních analýzách přecházet, tzv. dlouhá (autonomie, věk) vs. široká data (autonomie v 18, v 19, ..., ve 28 letech). V případě dat z projektu Cesty do dospělosti je problém s tím, že variabilita věků, v nichž byla autonomie měřena je velmi vysoká, a to jak díky možnosti časově flexibilně odpovídat na jednotlivé dotazníky, tak díky zrychlenému longitudinálnímu designu s různě starými účastníky na začátku studie. Pro LGC analýzu je tedy potřeba se rozhodnout, do jakých věkových intervalů individuální měření agregovat a jak. Užší intervaly znamenají více chybějících hodnot a méně agregace, širší méně chybějících hodnot, které jsou ovšem vykoupeny větší mírou agregace jednotlivých měření, což znamená určitou ztrátu dat. S ohledem na množství měření jsem se rozhodl data agregovat do ročních věkových intervalů (17,5-18,49; 18,5-19,49...) a jako agregační funkci použít průměr. To znamená, že pokud někdo během jednoho ročního intervalu vyplnil tři dotazníky projektu Cesty do dospělosti, průměr těchto tří hodnot tvoří jeho hodnotu autonomie pro daný věk (roční věkový interval). Tato míra agregace znamená, že prožívaná autonomie je zachycena 11 proměnnými (autonomie v 18, v 19, ..., ve 28). Volba

⁷ Jak je uvedeno v poznámce výše, Mplus umožňuje model vytvořit i bez této transformace. S použitím TSCORES lze LGCM model specifikovat tak, že náboje směrnice na manifestních měřeních jsou individuálně nastaveny podle přesných věků, v nichž se měření odehrálo. Výsledky jsou podobné jako u GCM v SPSS a na model lze pak aplikovat mixture analýzu. Nevýhodou proti klasickému LGCM je minimální množství diagnostických informací v tomto režimu (jen informační kritéria). Ani reziduální matice není součástí výstupu. Výpočty ztelně pomalejší a LMR ani BLRT nejsou v tomto režimu dostupné.

půlročního věkového intervalu by znamenala zdvojnásobení počtu proměnných a značný nárůst procenta chybějících hodnot v datové matici. Počet proměnných je zde důležitým faktorem, protože LGC model používá jako vstupní data samozřejmě také kovariance mezi jednotlivými měřeními, což při nárůstu měření z 11 na 22 znamená nárůst počtu kovariancí z 55 na 231.

Kvůli agregaci dat také klesl počet jednotlivých měření, s nimiž může pro jednotlivého účastníka model pracovat. S ohledem na tuto redukci jsem LGC modely odhadoval na datech s platnými hodnotami autonomie v minimálně třech věcích. Délka trvání studie (5 let) pak stanovuje maximum na 5 hodnotách. Popisné statistiky takto vytvořených proměnných jsou prezentovány v tabulkách IV.2 a IV.3. Hodnoty autonomie jsou stejně jako předchozích modelů umocněny na druhou. V souladu s GCM modely popsány výše průměrné hodnoty autonomie s věkem rostou. Postupně však také mírně rostou směrodatné odchylky. Pro většinu kovariancí platí, že čím vzdálenější v čase měření jsou, tím je jejich vztah slabší.

Tabulka IV.2. Průměry a směrodatné odchylky kvadratické autonomie pro věky 18-28 let

Věk	18	19	20	21	22	23	24	25	26	27	28
M	24,3	25,3	26,9	27,6	28,8	30,1	30,9	32,0	32,0	32,9	32,1
SD	7,53	8,04	7,86	8,44	8,41	8,83	8,78	9,14	9,39	8,61	9,82

Tabulka IV.3. Kovarianční matice opakovaných měření kvadratické autonomie pro věky 18-28 let

	18	19	20	21	22	23	24	25	26	27	28
AUTSQ18	56,8										
AUTSQ19	44,3	64,6									
AUTSQ20	33,8	39,2	61,7								
AUTSQ21	27,3	31,6	47,2	71,2							
AUTSQ22	36,9	29,4	41,2	53,0	70,8						
AUTSQ23	30,6	36,0	43,9	48,2	57,1	78,0					
AUTSQ24	18,1	13,5	34,6	42,5	51,0	58,8	77,1				
AUTSQ25	15,9	13,1	26,2	40,3	46,2	52,4	60,5	83,5			
AUTSQ26	15,5	11,9	25,3	35,5	43,8	49,5	58,3	65,4	88,2		
AUTSQ27	12,3	9,6	17,6	25,7	32,5	34,5	37,7	47,1	66,2	74,2	
AUTSQ28	4,2	3,3	2,6	5,9	9,7	8,7	4,9	15,3	48,7	59,2	96,4

Poznámka. Kovariance odhadnuté na základě FIML modelu v Mplus. Kovariance uvedené kurzívou jsou odhady, pro které nejsou na použitém vzorku žádná data (tj. rozestup věků je větší než 5 let).

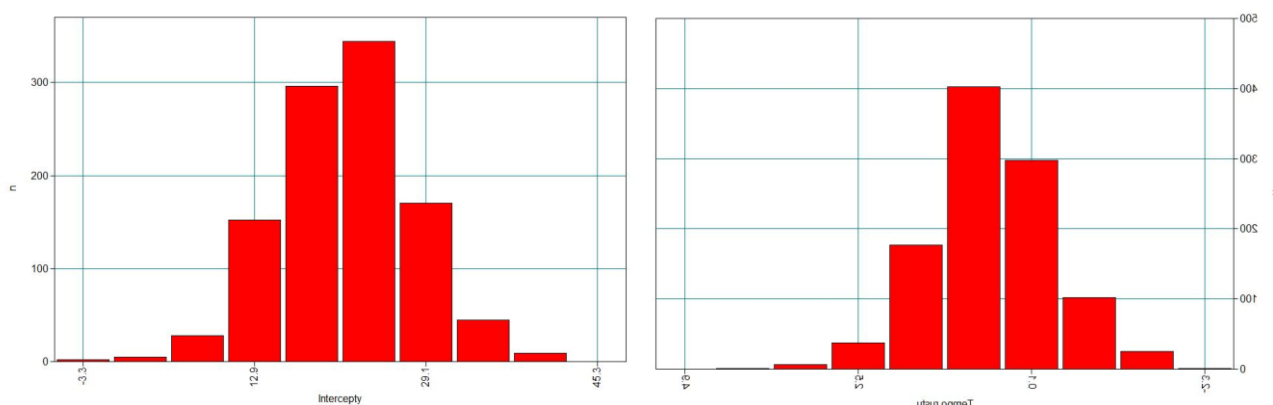
LGC MODELÝ AUTONOMIE

LGCM přístup je oproti GCM o něco flexibilnější ve specifikaci modelů, a tak v následujících pasážích představím vedle modelu lineárního růstu a kvadratického růstu také ještě několik příbuzných modelů. Parametry všech modelů jsou zde odhadovány v Mplus 7.11 (Muthén & Muthén, 2013) kritériem maximální věrohodnosti. Vzhledem k designem danému chybění dat je použita metoda full-information maximum likelihood.

Referenčním nulovým modelem je zde model konstantního průměru. Má jednu latentní proměnnou reprezentující výchozí hodnotu autonomie ve věku 18 let (průsečík). Tato latentní proměnná má nulový rozptyl. Reziduální rozptyly jsou volně odhadovatelné, jejich kovariance jsou fixované na nulu. To je podle Kennyho (2018, též McArdle, Nesselrode, 2014) vhodnější nulový model pro výpočet relativních indexů shody modelu s daty (např. CFI). V nulovém modelu je tedy 12 odhadovaných parametrů - 11 rozptylů a jeden průměr (průsečík). Samotné parametry tohoto modelu nejsou zajímavé, a proto nejsou v tabulkách prezentovány. Ukazatele shody tohoto modelu s daty, s nimiž budou následující modely srovnávány, jsou $\chi^2(49) = 3094$, AIC = 32192, BIC = 32252, SSABIC = 32213.

Model lineárního růstu s konstantními rezidui (M1_0) by měl být přímo srovnatelný s GCM modelem popsáným výše. Vývoj je v něm popsán dvěma latentními proměnnými - latentním průsečíkem (počáteční hodnoty) a latentní směrnicí (tempo růstu). Latentní průsečík má náboje na všech manifestních proměnných (autonomie v jednotlivých věcích) fixované na 1. Model předpokládá, že má normální rozložení s volně odhadovaným průměrem i rozptylem. Jeho průměr reprezentuje průměrnou hodnotu autonomie v 18 letech a rozptyl variabilitu odhadované počáteční úrovně autonomie mezi jednotlivci. Latentní směrnice má náboje na jednotlivých manifestních proměnných zafixované v pevných jednotkových intervalech od 0 (autonomie ve věku 18 let) do 10 (autonomie ve věku 28 let). Model předpokládá, že latentní směrnice má normální rozložení s volně odhadovaným průměrem i rozptylem. Jeho průměr reprezentuje průměrný roční přírůstek prožívané autonomie a rozptyl variabilitu ročních přírůstků mezi jednotlivci. Porovnáme-li parametry tohoto modelu s GCM modelem lineárního růstu, průměrné hodnoty jejich růstových parametrů jsou velmi podobné - z počátečních asi 24 (na škále autonomie umocněné na druhou 1-49) přibývá každým rokem asi 1 bod.

Obrázek IV.5. Rozložení průsečíků a směrnic v lineárním LGC modelu (M1_0)



Stochastická část modelu se předvídatelně liší. Všechny odhadované rozptyly a kovariance jsou nižší, což je způsobeno agregací dat do ročních věkových intervalů – část náhodného rozptylu se tím odstraní. Přesto zůstávají rozptyly průsečíků i směrnic poměrně velké a statisticky významně odlišné od 0. Směrodatná odchylka průsečíků je 8,7 a směrnic 1,4. Rozložení individuálních hodnot průsečíků a směrnic je na obrázku IV.5. Zůstává negativní korelace mezi průsečíky a směrnicemi ($r = -0,6$); čím vyšší je počáteční hodnota autonomie, tím pomalejší je růst.

Ukazatele shody modelu s daty založené na srovnávání modelu lineárního růstu (M1_0) s nulovým modelem (CFI, RMSEA) ukazují na dobrou shodu modelu s daty. Naopak absolutní ukazatele shody modelu s daty poukazují na to, že modelem implikovaná kovarianční matice se od empirické kovarianční matice liší ($\chi^2(55) = 192, p < 0,001, SRMR = 0,256$). Problém zde spočívá v tom, že kvůli zrychleně-longitudinálnímu modelu jsou maximálně pětileté individuální přímky v modelu extrapolované na desetileté rozpětí. Touto extrapolací se individuální růstové křivky dostávají mimo omezené rozpětí škály autonomie (viz Obrázek IV.3). Kvůli tomu model predikuje značně vyšší rozptyly, než jsou v pozorované kovarianční matici, zejména na okrajích věkového rozpětí. Částečně jde také o důsledek nevhodnosti lineárního modelu změny vzhledem k použité škále. V níže prezentovaných kvadratických modelech se tento problém zmenší.

Výše uvedený lineární LGC model s konstantní hodnotou reziduálního rozptylu napříč věky není jediným lineárním modelem, který lze specifikovat. Výchozím chováním Mplus je v této třídě modelů umožnit reziduálním rozptylům v jednotlivých věcích, aby se od sebe lišily. To znamená, že předpokládáme, že v různých věcích je hodnota autonomie implikovaná individuální růstovou křivkou různá, tj. náhodný rozptyl kolem růstové křivky se může měnit. Parametry lineárního modelu, který se od výše uvedeného M1_0 liší pouze heteroskedastickými rezidui (M1_1) jsou uvedeny v tabulce IV.4. Heteroskedastická rezidua spotřebují 10 deset parametrů a s ohledem na to není tento model o mnoho lepší než model s homoskedastickými rezidui, byť jde podle LRT testu o statisticky významné zlepšení (rozdíl $\chi^2(10) = 23, p = 0,011$). Většina odhadovaných reziduálních rozptylů se pohybuje od 15,1 do 22,7. Výjimku tvoří věk 28 let, kde je reziduální rozptyl 46,1, a věk 19 let, kde je reziduální rozptyl 28,7. Ve věku 28 let již je v datech projektu Cesty do dospělosti velmi málo respondentů, takže je rozptyl zatížen větší chybou odhadu. Zároveň je zde největší rozdíl mezi pozorovanými průměry a průměry predikovanými modelem, což ukazuje na meze lineárního růstového modelu.

V předchozích modelech byla rezidua napříč věky nezávislá, což odpovídá předpokladu maturace měřené charakteristiky, jejíž okamžité měření je zatíženo náhodnými vlivy (reprezentovanými rezidui). Korelace mezi opakovanými měřeními autonomie je pak pouze díky stabilně rostoucím hodnotám autonomie v modelu reprezentovanými latentním průsečíkem a směrnicí. Pokud bychom chtěli zohlednit také dlouhodobější situační vlivy, mohli bychom umožnit rezidui, aby korelovala. Tím vyjadřujeme, že naměřená hodnota autonomie ve věku T je ovlivněna nejen skutečnou hodnotou prožívané autonomie a náhodnými vlivy, ale také vlivy, které byly přítomny i předchozím, či předchozích měřeních. Je na zvážení, zda takový vliv reprezentují pouze korelace měření v sousedních věcích (autokorelace 1. řádu) nebo autokorelace vyšších řádů. Vzhledem k tomu, že estimační algoritmus rozhoduje o tom, jak velká část rozptylu měřené charakteristiky bude považována za rozptyl stabilně vyvíjejícího se rysu (*true variance*), a jaká část bude ponechána pro reziduální kovariance, jde o rozhodnutí, které má velký vliv na stochastickou část modelu. Korelace rezidui navíc mohou být v čase stále stejné nebo se mohou měnit, takže je zde velké množství voleb.

Tabulka IV.4. Parametry, odvozené statistiky a ukazatele shody LGC modelů.

	Lineární růst Konst. rezidua (M1_0)	Lineární růst heterosked. rezidua (M1_1)	Lineární růst - korelovaná rezidua (M1cor)	Lineární růst- Konstantní reliabilita (M1_GW)	Kvadratický růst konstantní rezidua (M2_0)	Kvadratický růst heterosked. rezidua (M2_1)	Kvadratický růst korelovaná rezidua (M2cor)
<i>Růstové parametry</i>							
Počáteční hodnota	24,6	24,6	24,7	24,6	23,6	23,6	23,7
(Věk-18)	1,05	1,05	1,03	1,06	1,56	1,55	1,52
(Věk-18) ²					-0,057*	-0,056*	-0,054*
<i>Rozptyly</i>							
Rezidua	19,8	15,1 - 46,1	18,3 - 64,7	17,8 - 38,9	18,4	7,3 - 30,0	22,3
Počáteční hodnota (I)	76,9	76,6	52,6	81,0	57,7	54,8	41,1
(Věk-18)	2,0	2,0	1,2	1,8	9,2	8,9	5,3
(Věk-18) ²					0,08	0,08	0,05*
COV (I; (Věk-18))	-7,5	-7,4	-3,4	-7,3	-10,0*	-9,0*	-3,3*
COV (I; (Věk-18) ²)					0,22*	0,17*	-0,23*
COV ((Věk-18); (Věk-18) ²)					-0,77	-0,75	-0,41*
Korelace reziduí	-	-	0,14 - 0,35	-	-	-	0,20
<i>Odvozené statistiky</i>							
r (I; (Věk-18))	-0,60	-0,60	-0,43	-0,61	-0,44	-0,41	-0,22*
r (I; (Věk-18) ²)					0,10*	0,08*	-0,17*
r ((Věk-18); (Věk-18) ²)					-0,88	-0,88	-0,84
<i>Ukazatele shody s daty</i>							
Chi2	192,0	169,0	75,5	196,7	106,7	83,8	65,8
df	55	45	35	55	51	41	50
Volných parametrů	6	16	26	6	10	20	11
RMSEA	0,042	0,051	0,033	0,050	0,032	0,032	0,017
CFI	0,955	0,959	0,987	0,953	0,982	0,986	0,995
SRMR	0,256	0,259	0,221	0,271	0,174	0,157	0,130
AIC	29279	29275	29202	29283	29201	29198	29162
BIC	29309	29354	29331	29313	29251	29298	29217
SSABIC	29290	29303	29249	29294	29219	29234	29182
<i>Test modelů</i>		M1_0 vs M1_1	M1_1 vs M1cor		M2_0 vs M1_0	M2_0 vs M2_1	M2_0 vs M2cor
rozdíl -2LL		23,0	116,5		85,3	22,9	40,9
rozdíl df		10	10		4	10	1
p		0,011	<0,001		<0,001	0,011	<0,001

Poznámka. Nemá-li parametr označen *, popř. není-li uvedeno jinak, $p < 0,01$.

Model M1cor prezentovaný coby příklad v tabulce IV.4 zahrnuje autokorelace reziduí 1. řádu, které se mohou v čase měnit spolu s heteroskedastickými rezidui. V souladu s předpoklady se parametry průměrné růstové křivky zařazením autokorelací nezměnily, zatímco jejich variabilita byla odhadnuta jako nižší. Reziduální rozptyly jsou naopak mírně vyšší. Korelace mezi nimi se pohybují mezi 0,14 a 0,35 s výjimkou korelace reziduí mezi 27. a 28. rokem, která byla odhadnuta na 0,46. Vzorec korelací je post hoc smysluplný. Vyšší hodnota 0,35 mezi 18. a 19. rokem, kdy je ještě většina mladých lidí v našem vzorku na střední škole, klesá na 0,14 mezi 19. a 20. rokem, kdy se životní podmínky většiny mladých lidí mění, ať již změnou školy, nástupem do práce apod. Mezi věky 20 a 23 jsou reziduální korelace blízké hodnotě 0,3, aby pak postupně klesaly k hodnotě 0,15. Vysokou hodnotu korelace mezi 27. a 28. rokem lze přičíst malé velikosti vzorku spojené s vysokou chybou odhadu. Ta je u kovariancí reziduí poměrně vysoká, pouze autokorelace mezi věky 20 až 24 jsou statisticky významné na 1% hladině. I když autokorelace nejsou nijak extrémně vysoké, jejich zařazení do modelu znatelně zlepšilo všechny ukazatele shody modelu s daty, což potvrzuje i LRT test modelu M1cor proti M1_1. Pouze BIC výrazně penalizující za zvýšený počet volných parametrů v modelu by preferoval model výchozí model M1_0. Rychle rostoucí počet parametrů je důležitým aspektem v rozhodování o výchozím modelu pro GMM, protože analýza směsí se při rostoucím počtu náhodných parametrů, jejichž variabilita je modelována jako směs, velmi rychle dostává do problému s konvergencí a identifikací modelu.

Pro zajímavost prezentuji ještě jednu variantu lineárního modelu, kde jsou rezidua podle Grimma a Widamana (2010) specifikována tak, aby byla napříč věky konstantní reliabilita růstové křivky, tedy podíl rozptylu autonomie v daném věku vysvětlený modelem. Tato reliabilita není stanovena dopředu, ale nepřímo se odhaduje jako jeden z parametrů modelu. V případě lineárního modelu autonomie byl tento parametr odhadnutý na 0,74, což znamená, že v každém věku byl reziduální rozptyl modelem nastaven na 26 % rozptylu kvadratické autonomie v daném věku. Reziduální rozptyly tak kopírují měnící se variabilitu během času a nabývají hodnot od 17,8 do 38,9. Růstové parametry tím stále nejsou ovlivněny a jejich variabilita se příliš neliší od modelu s konstantními rezidui. Přesto je shoda tohoto modelu s daty ve všech ukazatelích mírně horší.

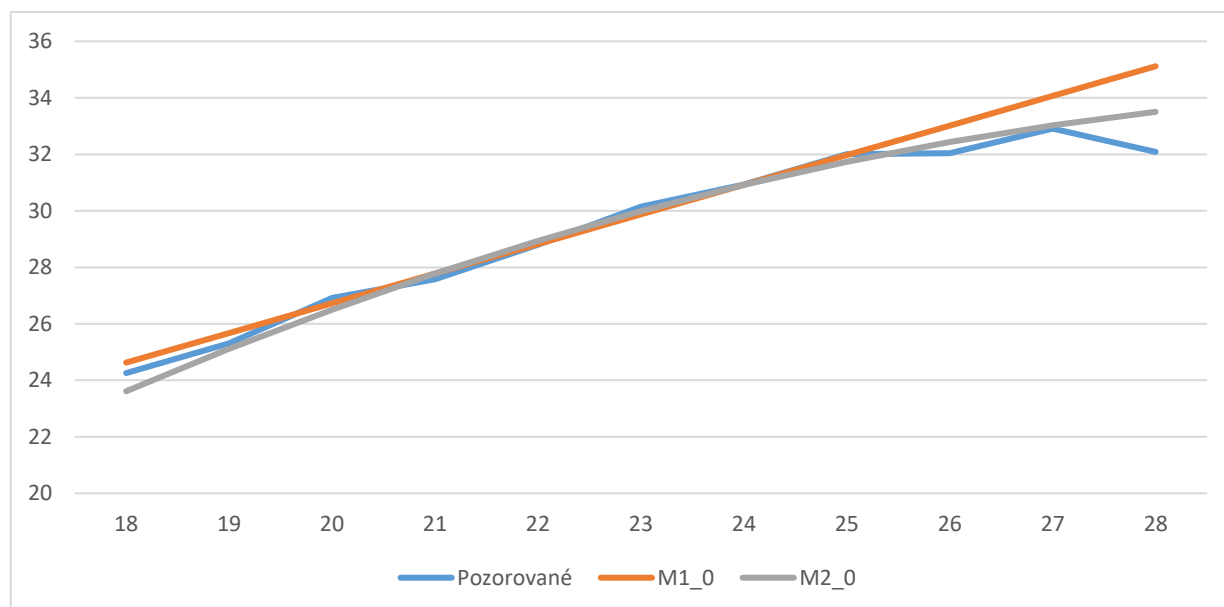
Dalšími lineárními modely, které by bylo možné použít pro modelování vývoje autonomie, jsou level-and-shape popř. latent-base modely, které jsou lineárními pouze z formálního hlediska. V těchto modelech je tempo růstu zafixováno náboji latentní směrnice pouze na dvou či více měřeních (věcích) a ostatní náboje se odhadují jako volné parametry modelu. V případě autonomie by tedy bylo možné nastavit náboj latentní směrnice ve věku 18 na 0 a ve věku 28 na 10 a v ostatních věcích nechat náboje volně odhadnutelné. Pokud by jejich hodnoty byly odhadnuty blízko hodnot 1,2,...,8,9 mohli bychom považovat růst za lineární, ale jeho tvar může být vlastně libovolný. Také je možné použít tzv. piecewise modely, kdy jsou odhadováno více latentních směrnic, každá pro jiné věkové rozmezí. V případě dat z Cest do dospělosti tyto uvolněné modely nedávají dobré výsledky, kvůli designem studie implikovanému chybění dat, kdy žádný účastník nemá data pro celé desetileté rozpětí. Růstové parametry jsou odhadovány s velkou směrodatnou chybou a od lineárního růstu se příliš nevzdalují. Pro vyrovnání se s chybějícími daty je naopak výhodná jednoduchá polynomičká forma růstu překlenující celé věkové období s využitím minima parametrů.

Nedostatky výše uvedených lineárních modelů ukazují, že by bylo vhodné je srovnat s kvadratickým růstovým modelem. LGC model kvadratického růstu, který je přímo srovnatelný s GCM modelem kvadratického růstu prezentovaným výše, je model s konstantními rezidui. Oproti analogickému lineárnímu modelu přibývá latentní kvadratická

směrnice, jejíž náboje jsou fixovány na hodnotu druhých mocnin nábojů latentní (lineární) směrnice. Její význam je také analogický – její průměrná hodnota je kvadratickým koeficientem růstu průměrné růstové trajektorie a její rozptyl reprezentuje variabilitu těchto kvadratických růstových koeficientů mezi jednotlivci. Protože jsou nyní v modelu tři latentní růstové proměnné, odhadují se také tři kovariance mezi nimi. Kromě kovariance mezi průsečíkem a lineárním růstem je to kovariance mezi průsečíkem a kvadratickým růstem a kovariance mezi lineárním a kvadratickým růstem. Ačkoli je obtížné si představit, vizualizovat společný efekt všech těchto tří kovariancí, nedoporučuje se je fixovat. Kvadratický model tak vyžaduje o čtyři odhadované parametry více.

Základní LGC model kvadratického růstu s homoskedastickými rezidui je shrnut v tabulce IV.4 (M2_0). Růstové parametry se příliš neliší od lineárního modelu, pouze je mírně modifikují. Počáteční úroveň percipované autonomie je pro průměrnou růstovou křivku mírně nižší, což kompenzuje o necelého půl bodu rychlejší lineární roční růst. Tempo růstu se však díky kvadratickému koeficientu postupně zpomaluje a na konci modelovaného rozpětí je už jen asi půl bodu ročně. Pro lepší představu jsou v obrázku IV.6 zobrazeny pozorované průměry a průměrné trajektorie implikované lineárním a kvadratickým modelem.

Obrázek IV.6. Průměrné trajektorie implikované LGC růstovými modely

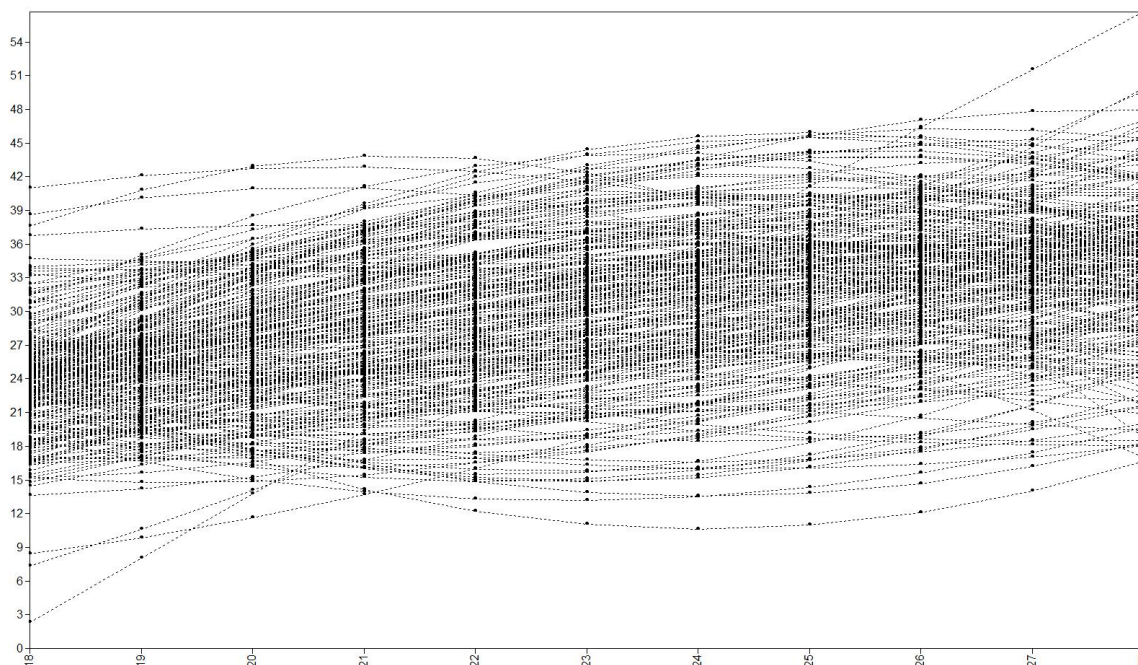


Pozn. Zobrazeny jsou modely s konstantními nekorelovanými rezidui. Ostatní varianty modelů se v průměrné vývojové trajektorii téměř neliší.

Průměrná hodnota latentní kvadratické směrnice sice není statisticky významně odlišná od 0 ($z = -2,4$, $p = 0,015$), ovšem nejde o hypotézu, která by nás u tohoto modelu zajímala – důležitější je to, že rozptyl tohoto růstového parametru (i zbývajících dvou) je statisticky významně nenulový. Lze si například představit model, v němž je průměru nulový růst, ale polovina populace roste a polovina klesá. Zde počáteční hodnoty variují se směrodatnou odchylkou 7,6, lineární růstový parametr se směrodatnou odchylkou 3,0 a kvadratický se směrodatnou odchylkou 0,28. Oproti lineárnímu modelu je v kvadratickém modelu prostor pro podstatně rychlejší změny prožívané autonomie (tam byly SD lineárního růstu pouze 1,1-1,4), které mohou být kompenzovány kvadratickým členem. Mluvit o kompenzaci je vskutku na místě, protože korelace mezi individuálními hodnotami lineárního a kvadratického parametru je téměř -0,9. Rychlejší počáteční růst je tak téměř plně vyvážen intenzivnějším zbrzděním růstu kvadratickým růstovým parametrem. Představu zde ještě dále komplikuje

slabá pozitivní korelace mezi počáteční hodnotou prožívané autonomie a lineárním růstovým parametrem. Pro lepší představu je na obrázku IV.7 zobrazeno 300 náhodně vybraných predikovaných vývojových trajektorií, z nichž je patrné, jaké různé trajektorie v sobě model zahrnuje.

Obrázek IV.7. Náhodný výběr 300 predikovaných individuálních trajektorií v kvadratickém LGC modelu



Stejně jako u lineárního modelu platí, že jeho růstové parametry jsou velmi podobné kvadratickému GCM modelu prezentovanému výše a jeho rozptylové parametry jsou znatelně nižší kvůli agregaci dat.

Shoda kvadratického LGC modelu s daty je oproti lineárnímu modelu lepší ($\chi^2(4) = 85,3$, $p < 0,001$). Jednotlivé ukazatele shody s daty jsou také všechny znatelně lepší, včetně informačních kritérií penalizujících za nárůst počtu parametrů. Zejména SRMR znatelně poklesl, byť stále zůstává v hodnotách, které jsou považovány za známku špatné shody modelu s daty. Zdrojem této hodnoty jsou stále obtížně odhadnutelné kovariance mezi věky, které jsou od sebe vzdálené více než 5 let.

Varianta kvadratického LGC modelu s heteroskedastickými rezidui má jen marginálně lepší shodu modelu s daty než model s konstantními rezidui. Je zajímavé, že vedle nezměněných růstových parametrů došlo k drobnému poklesu odhadnutého rozptylu latentního průsečíku a latentní lineární směrnice. Deset parametrů navíc by komplikovalo mixture analýzy, a proto pokládám za vhodnější model s konstantními rezidui.

Další variantou kvadratického modelu je model autokorelovanými rezidui. S ohledem na nárůst počtu parametrů a na to, že teorie neukazuje jasně na to, že by se rezidua a jejich autokorelace měly postupně měnit, prezentuji zde model homoskedastickými autokorelovanými rezidui s konstantní mírou autokorelace (M2cor). Tento model vykazuje značné zlepšení všech ukazatelů shody modelu s daty, a to s jediným parametrem navíc (LRT: $\chi^2(1) = 40,9$, $p < 0,001$). S výjimkou SRMR jsou všechny ukazatele shody modelu s daty uspokojivé. Růstové parametry zůstaly stejné jako v modelu bez autokorelace rezidui. V důsledku zařazení autokorelace se mírně navýšil reziduální rozptyl a spolu s ním poklesly

odhadnuté rozptyly latentních růstových parametrů. Odhad rozptylu kvadratické latentní směrnice se tak dostal na hranici signifikance ($z = 2,2$, $p = 0,030$). S tím klesly také hodnoty kovariance mezi latentními růstovými parametry; pouze kovariance mezi latentním lineárním a latentním kvadratickým růstovým parametrem je statisticky signifikantní ($z = 2,0$, $p = 0,041$). To znamená, že tempo růstu prožívané autonomie nezávisí na počáteční hodnotě růstu. Tento model tak představuje interpretačně i parametricky velmi dobrý poměr mezi komplexitou modelu a shodou s daty. V tomto smyslu jde o model vhodný pro následnou mixture analýzu, která parametry s přibývajícím počtem latentních tříd znásobuje.

Volba nepodmíněného modelu růstových křivek je prvním krokem v analýze heterogenity růstových trajektorií. Výše představené latentní růstové modely dokladují, že ačkoliv se různě specifikované modely nemusí velmi lišit odhadovanou průměrnou trajektorií, liší se odhadovanou variabilitou růstových parametrů, a tedy i odhadem toho, jak velká je heterogenita, již bychom chtěli dále analyzovat. Uvedené růstové modely se značně liší i počtem volných parametrů (od 6 do 26), což je při volbě modelu vhodného pro exploraci latentních tříd důležitý faktor. Podívejme se nyní, jak se různé modely latentních růstových křivek chovají při aplikaci modelu latentních růstových tříd.

Modely latentních růstových tříd autonomie

Modely latentních růstových tříd (*latent class growth analysis, LCGA*) se heterogenitu růstových trajektorií usilují zachytit pomocí určitého množství tříd, kategorií, v rámci kterých jsou růstové parametry konstantní, pro všechny členy třídy shodné. Veškeré odchylky od třídně-specifické vývojové trajektorie jsou tak pokládány za náhodné chyby a variabilita se přesouvá do reziduálního rozptylu. Odhadují se tak pouze růstové parametry pro jednotlivé třídy, velikosti tříd a reziduální rozptyl. Základní analytickou strategií je zde postupné odhadování modelů s vyšším počtem tříd, dokud budou třídy smysluplně velké, separované a model s vyšším počtem tříd bude lépe odpovídat datům při zohlednění nárůstu počtu parametrů. Počet tříd však není to jediné, o čem je nutné se při specifikaci modelu rozhodnout. Na základě téhož LGC modelu lze specifikovat třídy, které se liší pouze růstovými parametry, nebo i reziduální strukturou růstového modelu.

LCGA LINEÁRNÍHO MODELU S HOMOSKEDASTICKÝMU REZIDUI

Podívejme se nyní na nejjednodušší z výše popsaných modelů růstu prožívaní autonomie mezi věky 18 a 28 – model latentního lineárního růstu s konstantními rezidui. V rámci tohoto modelu jsou rozdíly mezi jednotlivci popsány rozptylem počátečních úrovní prožívané autonomie, rozptylem směrnic a jejich kovariancí. Model latentních růstových tříd se dvěma třídami tak popisuje veškerou variabilitu individuálních růstových trajektorií pomocí dvou růstových parametrů pro každou třídu (počáteční úroveň a tempo růstu/poklesu), relativní četnosti třídy (celkem počet tříd - 1) a reziduálního rozptylu.

V rámci specifikování LCGA modelu je možné reziduální strukturu odhadovat pro každou třídu zvlášť, nebo snížit počet odhadovaných parametrů omezením reziduí tak, aby byly napříč třídami stejné.

Tabulky IV.5 a IV.6 shrnují parametry série LCGA modelů vycházejících z lineárního latentního modelu s konstantními rezidui shodnými napříč latentními třídami (M1_0). Pro model se dvěma třídami je odhadnutá velikost tříd v populaci 54 % a 46 % populace. Model tedy dělí populaci přibližně na poloviny, přičemž v té první začíná lineární růst na poměrně vysoké hodnotě 30,1 a pokračuje tempem 1,1 bodu za rok. V druhé třídě začíná růst níže, na

19,4, a s 0,8 body přírůstku ročně je mírně pomalejší. Reziduální rozptyl byl v obou skupinách odhadován společně a má vysokou hodnotu 38,6 bodu. Pro členy modelem definovaných subpopulací tak platí, že jejich individuální hodnoty se od třídní růstové trajektorie náhodně odchylují se směrodatnou odchylkou 6,2 bodu (na kvadratické škále autonomie od 1 do 49). Není zde však nijak rozlišeno to, nakolik je tento rozptyl utvářen kolísáním individuálních hodnot kolem třídní růstové trajektorie, a nakolik tím, že se celé individuální trajektorie odklánějí od třídní růstové trajektorie. Nepřekvapí tedy, že LCGA model se dvěma kategoriemi odpovídá datům podstatně hůře než výchozí LCG model, ať již jej soudíme podle velikosti reziduí nebo podle informačních kritérií. I když v tabulce IV.5 jsou uvedeny také hodnoty testů LMR a BLRT, ty nejsou příliš informativní, protože srovnávají LCGA model se dvěma třídami s LCGA modelem s jednou třídou (a ne s LGC modelem). Jejich signifikace je triviální konsekvencí toho, že růstové parametry jsou vskutku heterogenní.

Tabulka IV.5. Ukazatele shody modelů s daty pro LCGA modely vycházející z M1_0

M1_0	par	LL	AIC	BIC	SSABIC	Entropie	LMR	LMR p	BLRT	BLRT p
LGC	6		29279	29309	29290					
C2	6	-15141,1	30294	30324	30305	0,82	1639,0	<0,0001	1717,6 (3)	<0,0001
C3	9	-14862,2	29742	29787	29758	0,81	532,2	<0,0001	557,7 (3)	<0,0001
C4	12	-14753,4	29531	29590	29552	0,77	207,6	0,001	217,6 (3)	<0,0001
C5	15	-14721,2	29472	29547	29499	0,75	61,5	0,088	64,5 (3)	<0,0001
C6	18	-14696,9	29430	29519	29462	0,70	46,4	0,263	48,7 (3)	<0,0001

Zaměříme-li se na separaci tříd, hodnota entropie je u modelu se dvěma třídami uspokojivě vysoká (0,82) a odráží to, že průměrná klasifikační pravděpodobnost pro třídu, do níž jednotlivec nenáleží, je pro obě třídy menší než 6 %. Pokud by tedy skutečně v populaci existovaly subpopulace odpovídající třídám v tomto modelu, pravděpodobnost špatné klasifikace (určení příslušnosti k subpopulaci podle naměřených hodnot prožívané autonomie) by byla v každé subpopulaci menší než 6 % (celkem by byla pravděpodobnost špatné klasifikace 5 %)

LCGA model se třemi a čtyřmi třídami jsou postupným zpřesněním modelu se dvěma třídami. Datům odpovídají lépe než předchozí model prizmatem všech informačních kritérií i testů LMR a BLRT. I navzdory rostoucímu počtu tříd, zůstávají třídy dostatečně separované, o čemž svědčí jen mírně klesající hodnota entropie. Oproti hierarchickým shlukovacím metodám se napříč modely mohou třídy libovolně lišit. Také jejich pořadí je čistě arbitrární, což je drobnost znepríjemňující srovnávání výsledků mezi modely navzájem. Například model se třemi třídami implikuje 3 subpopulace relativních četnostech 47 %, 22 % a 31 %. I když by se mohlo zdát, že druhé dvě třídy mohly vzniknout rozdělením první třídy ve dvoutřídním modelu (s relativní četností 54 %), z pohledu na růstové křivky je zřejmé, že to tak není. Proto je zajímavé u každého jednotlivého modelu sledovat, jak se složení tříd proměňuje s nárůstem jejich počtu. U dvoutřídního modelu se třídy lišily spíše jen počáteční úrovní autonomie, tempo růstu bylo v obou třídách podobné. U modelu se třemi třídami se první a třetí třída, které dohromady tvoří téměř 80 % populace, opět liší jen počáteční hodnotou prožívané autonomie. Ovšem druhá třída představuje subpopulaci s nízkou počáteční autonomií a polovičním tempem růstu. V modelu s čtyřmi třídami se také vyskytuje

třída s nízkou počáteční úrovní autonomie, a ještě nižším tempem růstu (třída č. 3), nejde však o totožnou subpopulaci jako v modelu se třemi třídami – v modelu se čtyřmi třídami je tato subpopulace odhadnuta mnohem menší, pouze 12 %. Zbývající třídy v modelu se čtyřmi třídami se opět liší jen počáteční úrovní autonomie. U modelů se 3 a 4 třídami je patrný také pokles reziduálního rozptylu s rostoucím počtem tříd.

Tabulka IV.6. Parametry LCGA modelů vycházejících z M1_0

Model		C2	C3	C4	C5	C6
<i>Relativní četnost</i>	C#1	0,54	0,47	0,37	0,28	0,22
	C#2	0,46	0,22	0,18	0,05	0,09
	C#3		0,31	0,12	0,23	0,31
	C#4			0,32	0,32	0,16
	C#5				0,11	0,17
	C#6					0,05
<i>Růstové parametry</i>						
C#1	Průsečík	30,11	23,47	26,50	19,71	31,50
	Směrnice	1,07	1,10	1,27	0,95	1,23
C#2	Průsečík	19,42	16,77	34,94	40,11	16,65
	Směrnice	0,80	0,55	1,04	0,62	-0,08
C#3	Průsečík		32,78	15,96	31,16	25,37
	Směrnice		1,07	0,26	1,25	1,26
C#4	Průsečík			20,34	25,13	14,72
	Směrnice			1,00	1,20	2,08
C#5	Průsečík				15,7	25
	Směrnice				0,22	-0,13
C#6	Průsečík					40,5
	Směrnice					0,58
Reziduální rozptyl		38,63	30,05	26,77	25,55	24,49

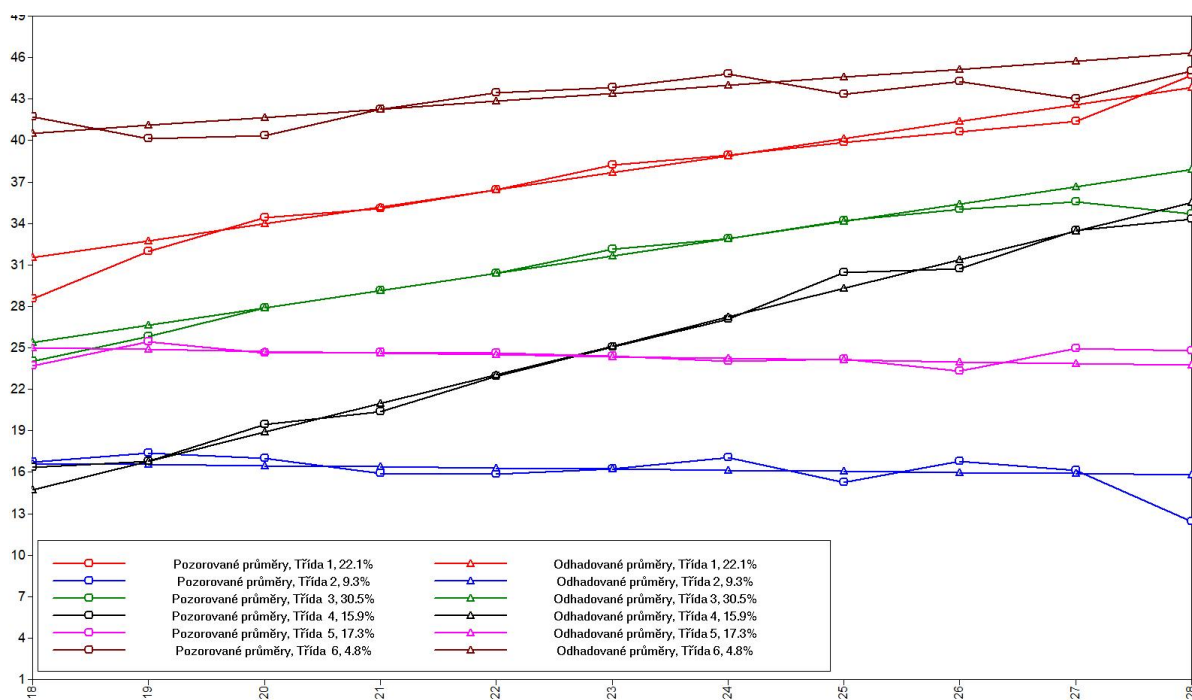
Model s pěti třídami do tohoto rozvíjejícího se vzorce přidává nový prvek. Vedle malé třídy s nízkou a nerostoucí autonomií (č. 5) vymezuje ještě menší třídu s velmi vysokou a pomaleji rostoucí autonomií (č. 2). Velikost této třídy model odhaduje na pouhých 5 %, což již s sebou přináší otázku, zda by opravdu mohlo jít o skutečnou subpopulaci, či jen o artefakt aktuálního vzorku. K těmto pochybám přispívá i to, že podle LMR testu není shoda modelu s 5 třídami signifikantně lepší než shoda modelu se 4 třídami. Tento test však Muthén a Asparouhov (Asparouhov & Muthén, 2012) podávají spíše jako výpočetně nenáročný orientační test. Všechny ostatní ukazatele, včetně BLRT testu model s 5 třídami preferují před modelem se 4 třídami. Třída s velmi malou relativní četností s sebou přináší i další nejistoty. Málo četná třída znamená nízkou apriorní pravděpodobnost klasifikace do této třídy, což

zvysuje pravdepodobnost spatne klasifikace clenů této hypotetizovane populace. Zde tento problem není až tak výrazný – u těch, kdo byli klasifikováni do této třídy, je pravdepodobnost 0,74, že skutečně pochází z této subpopulace. U ostatních tříd je tato pravdepodobnost 0,8 a vyšší. Je tedy pravdepodobné, že asi čtvrtina jednotlivců klasifikovaných do této třídy, je ve skutečnosti členy jiných tříd (zde specificky třídy č. 3).

LCGA lineární model se šesti třídami je podle informačních kritérií také lepší než předchozí model. LMR test opět není signifikantní (a zůstal by takovým i pro další modely s více třídami), BLRT (stále s 500 bootstrapovými vzorky z k-1 modelu) je stále s velkou rezervou signifikantní. Při šesti třídách již je pravdepodobné, že některé z nich budou malé. Zároveň se při takto vysokém počtu tříd již obtížně hledají teoretické důvody, proč by mohlo/mělo existovat takové množství teoreticky zdůvodněných subpopulací. Se šesti třídami spočívá smysl kategorizování spíše v zachycení heterogenity do kategorií a facilitaci person-centered teoretizování.

Pro snadnější interpretaci jsou tentokrát vývojové trajektorie v jednotlivých třídách zobrazeny v grafu, který produkuje Mplus (obrázek IV.8) spolu s pozorovanými průměry prožívané autonomie v jednotlivých věcích. Třídy 1, 3, 4, 6 jsou rostoucí a mírně konvergují k vyšším hodnotám prožívané autonomie a dohromady zahrnují tři čtvrtiny populace. Třídy 2 a 5 stagnují či mírně klesají z mírně podprůměrných (třída 5) či více podprůměrných (třída 2) počátečních hodnot prožívané autonomie. Na původní (neumocněné) škále autonomie od 1 do 7, kde hodnoty nad 4 znamenají souhlas s výroky značícími prožívanou autonomií to znamená, že třídy 1 a 3 popisují majoritní (>50%) rostoucí tendenci z hodnot kolem 5–6 k hodnotám kolem 6–6,5. Třída 5 kontrastuje s třídou 3, když ze stejné úrovně kolem 5 nestoupá. I hodnotu 5 však můžeme interpretovat jako převážné souhlasení s položkami škály autonomie, takže můžeme mluvit o jednotlivcích, kteří jsou zdrženlivě autonomní. Podobnou dvojici tvoří třídy 4 a 2, které obě začínají na hodnotách kolem 4 na původní škále pozorované autonomie, ale členové třídy 4 (16 %) se připojují k rostoucímu trendu, zatímco členům třídy 2 (9 %) prožívaná autonomie neroste.

Obrázek IV.8. Pozorované a odhadnuté průměry v LCGA lin. modelu se 6 třídami



Ukazatele shody modelu s daty až doposud stále stoupaly, a tak lze věřit, že i model se sedmi třídami by mohl pasovat ještě lépe. Kvůli absenci variability růstových křivek uvnitř tříd je navyšování počtu tříd jedinou možností, jak přesunout rozdíly mezi jednotlivci z reziduálního rozptylu do modelu. Vytrácí se ale smysl takového modelu. Z modelu se šesti třídami a modelu předcházejícího je zřejmé, že většina tříd vlastně jen kategoricky aproximuje rozložení růstových křivek. Je otázkou, k čemu by byla ještě jemnější aproximace. Z perspektivy identifikace existujících subpopulací je další navyšování počtu tříd také problematické, protože pro třídy s paralelními vývojovými křivkami lišícími se o jeden bod na původní sedmibodové škále autonomie se obtížně hledá kvalitativně odlišná interpretace. Z uvedených důvodů zde série LCGA modelů vycházejících z lineárního LGC modelu s konstantními rezidui končí.

Stojí za povšimnutí, že ani model se šesti třídami nepasuje na data tak dobře jako výchozí LGC model, který má navíc méně parametrů. Z hlediska parsimonie se tak lepším popisem heterogenity růstu jeví model se spojitým rozložením růstových parametrů, než model vnitřně růstově homogenních tříd.

LCGA MODELY S HETEROSKEDASTICKÝMI REZIDUI

Podobné výsledky skýtají LCGA modely vycházející z lineárního LGC modelu s heteroskedastickými rezidui. Ukazatele shody modelu s daty a parametry modelů jsou uvedeny v tabulkách IV.7 a IV.8. I zde i model se šesti třídami lépe pasuje na data než model s pěti třídami. Tři z těchto tříd jsou však odhadnuty jako poměrně malé (<10 %). Zajímavé je, že zbývající větší třídy neodrážejí zápornou kovarianci mezi počáteční úrovní prožívané autonomie a tempem růstu – růstové křivky jsou spíše paralelní než sbíhající se.

Tabulka IV.7. Ukazatele shody modelů s daty pro LCGA modely vycházející z M1_1

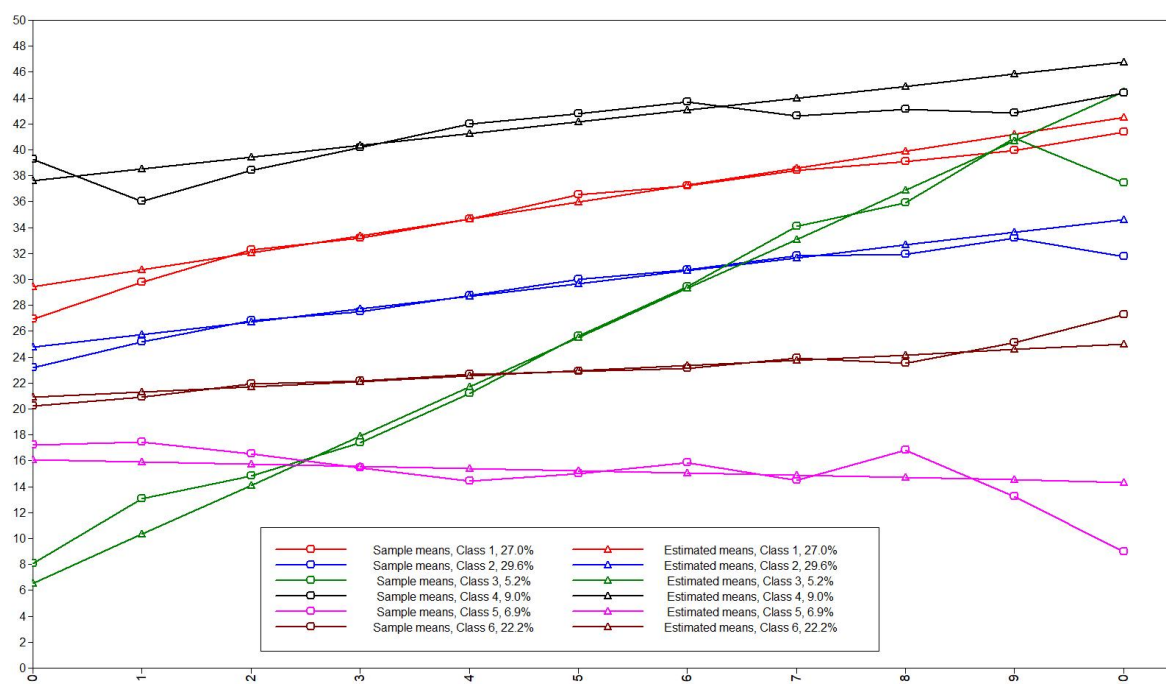
M1_1	par	LL	AIC	BIC	SSABIC	Entropie	LMR	LMR p	BLRT	BLRT p
LGC	16		29275	29354	29303					
C2	16	-15132,1	30296	30376	30325	0,83	1643,9	<0,0001	1722,7 (3)	<0,0001
C3	19	-14851,0	29740	29834	29774	0,81	536,4	0,000	562,1 (3)	<0,0001
C4	22	-14736,3	29517	29626	29556	0,78	219,0	0,006	229,5 (3)	<0,0001
C5	25	-14700,6	29451	29575	29496	0,75	68,1	0,462	71,4 (3)	<0,0001
C6	28	-14678,2	29412	29551	29462	0,74	42,8	0,353	44,8 (3)	<0,0001

Na M1cor a M1GW nyní LCGA analýzu nebudu aplikovat, protože je pravděpodobné, že bych dospěl k podobným výsledkům jako u modelů M1_0 a M1_1. Přejdu ke kvadratickým modelům, abych se podíval, jaké třídy lze nalézt v kvadratických vývojových trajektoriích. Začnu modelem kvadratického latentního růstu s konstantními rezidui – parametricky úsporný model, který maximalizuje variabilitu růstových parametrů.

Tabulka IV.8. Parametry LCGA modelů vycházejících z M1_1

Model		C2	C3	C4	C5	C6
<i>Relationí četnost</i>	C#1	0,53	0,21	0,13	0,23	0,27
	C#2	0,47	0,47	0,33	0,32	0,30
	C#3		0,31	0,37	0,06	0,05
	C#4			0,17	0,29	0,09
	C#5				0,10	0,07
	C#6					0,22
<i>Růstové parametry</i>						
C#1	Průsečík	30,19	16,80	15,97	19,18	29,43
	Směrnice	1,05	0,53	0,30	0,67	1,31
C#2	Průsečík	19,38	23,41	20,49	22,92	24,76
	Směrnice	0,81	1,12	1,03	1,22	0,99
C#3	Průsečík		32,82	26,79	14,67	6,52
	Směrnice		1,07	1,28	0,03	3,80
C#4	Průsečík			35,23	28,91	37,60
	Směrnice			1,05	1,31	0,91
C#5	Průsečík				36,84	16,10
	Směrnice				1,01	-0,18
C#6	Průsečík					20,91
	Směrnice					0,41
Reziduální rozptyl (min-max)		32-71	27-67	23-73	20-50	17-55

Obrázek IV.9. Pozorované a odhadnuté průměry v LCGA lin. modelu se 6 třídami



LCGA KVADRATICKÉ MODELY S HETEROSKEDASTICKÝMI REZIDUI NAPŘÍČ TŘÍDAMI

V modelech s konstantními i časově rozlišenými rezidui byla doposud rezidua stejná napříč třídami. Přitom, kdybychom hledali teoretické důvody, bylo by patrně snazší je najít pro shodu v čase než pro shodu mezi třídami. V řadě kontextů je plauzibilní představa, že určitá část populace má sledovanou charakteristiku poměrně stabilní a jiné zřetelně více kolísá, a to kolem téže vývojové trajektorie. Zvláště u charakteristik prožívání, temperamentových charakteristik jsou rozdíly ve stabilitě v čase součástí definice těchto konstruktů. Mixture modely umožňují specifikovat model tak, že se rezidua mohou mezi třídami lišit. Často to však znamená problémy s identifikací některých parametrů modelu a navýšení problémů s konvergencí. Metaforicky, model se musí rozhodovat, zda případy, které jsou vzdálenější od růstové trajektorie dané třídy, ponechat ve třídě a zvýšit její reziduální rozptyl, nebo z nich udělat samostatnou třídu. To prudce navyšuje paletu plauzibilních hodnot parametrů. Model třídně-specifickými rezidui lze vnímat jako dílčí krok směrem k růstovým mixture modelům. I když ještě neumožňuje přímo modelovat variabilitu růstových křivek uvnitř tříd, umožňuje ji alespoň rozlišit.

Následující LCGA modely specifikují postupně rostoucí počet latentních tříd kvadratických růstových křivek. Rozdíly mezi třídami spočívají nejen v rozdílných růstových parametrech – průsečíku, lineární a kvadratické směrnici – ale také v tom, že různé třídy mohou mít různá rezidua. Model se dvěma třídami tak má 9 volných parametrů: 3 růstové parametry pro každou třídu, konstantní reziduální rozptyl pro každou třídu a jednu relativní četnost třídy. S každou třídou přibývá dalších pět parametrů. Parametry shody modelu s daty se opět až do modelu se šesti třídami zlepšují. Zajímavě zde vychází LMR test, který již čtvrtou třídu považuje za redundantní, ale pro test modelu se 6 třídami vychází opět signifikantně.

Tabulka IV.9. Ukazatele shody modelů s daty pro LCGA modely vycházející z M2_0

M2_0	par	LL	AIC	BIC	SSABIC	Entropie	LMR	LMR p	BLRT (Δ par)	BLRT p
LGC	10		29201	29251	29219					
C2	9	-15130,5	30279	30323	30294	0,82	1681,7	<0,001	1730,1 (5)	<0,001
C3	14	-14847,4	29723	29792	29748	0,79	550,0	0,002	566,0 (5)	<0,001
C4	19	-14739,1	29516	29610	29550	0,76	210,5	0,091	216,5 (5)	<0,001
C5	24	-14690,7	29429	29548	29472	0,78	94,2	0,072	96,8 (5)	<0,001
C6	29	-14647,3	29352	29496	29404	0,76	84,3	0,032	86,7 (5)	<0,001

Pro modely se dvěma až čtyřmi třídami platí, že růstové trajektorie v třídách jsou přibližně paralelní, mírně konvergující. Vždy třída s nejnižší růstovou trajektorií má odhadnutý nejvyšší reziduální rozptyl. To může být artefaktem daným tím, že taková třída je nejdále od stropu škály. Může ale také vyjadřovat to, že střední hodnoty autonomie nejvíce fluktuují (stabilně nízké trajektorie se v datech příliš neobjevují). V modelu s 5 a 6 třídami se objevuje velmi malá třída (2 % populace), která má řádově menší reziduální rozptyl oproti ostatním třídám. Zde již je to zcela zřejmě artefakt stropu škály, protože tato třída má svou růstovou trajektorii těsně pod tímto maximem. Lze však uvažovat, že taková subpopulace

může existovat – lidé, kteří nemají (nebo nemohou mít) pochyby o své autonomii, a tak se jen zřídka vzdálí od maxima škály.

Tabulka IV.10. Parametry LCGA modelů vycházejících z M2_0

Model		C2	C3	C4	C5	C6
<i>Relativní</i>	C#1	0,49	0,44	0,34	0,13	0,02
<i>četnost</i>	C#2	0,51	0,30	0,34	0,32	0,08
	C#3		0,26	0,15	0,02	0,22
	C#4			0,17	0,21	0,12
	C#5				0,32	0,30
	C#6					0,26
<i>Růstové parametry</i>						
C#1	I	19,37	23,48	20,79	17,35	44,40
	S	0,97	1,37	1,07	-0,43	0,03
	Q	-0,02	-0,03	0,00	0,08	0,05
C#2	I	28,78	31,19	25,33	24,67	23,43
	S	1,88	2,00	2,02	1,91	0,538
	Q	-0,09	-0,10	-0,07	-0,07	0,042
C#3	I		17,62	17,28	44,45	30,81
	S		0,48	-0,24	0,02	2,234
	Q		0,02	0,07	0,05	-0,118
C#4	I			33,08	30,83	16,44
	S			2,10	2,31	0,013
	Q			-0,114	-0,13	0,027
C#5	I				20,14	24,34
	S				1,13	1,991
	Q				-0,013	-0,077
C#6	I					19,71
	S					1,302
	Q					-0,036
Reziduální rozptyly		42	26	28	33	6
		35	29	23	23	79
			38	34	6	24
				26	24	27
					28	16
						21

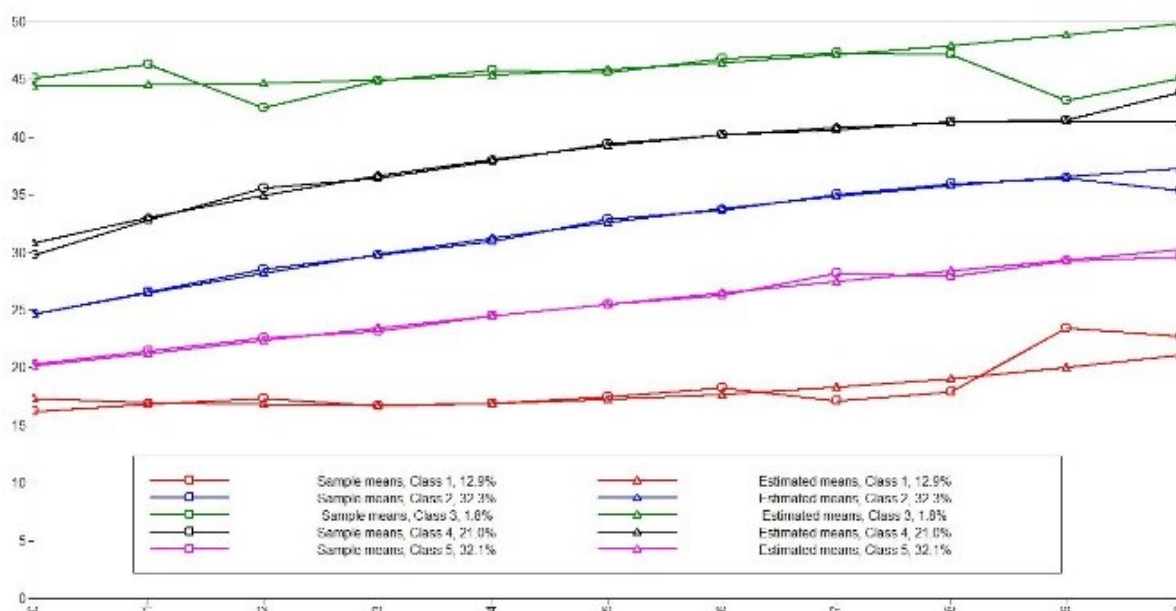
Poznámka. Červeně zvýrazněné jsou problematické parametry.

V modelu se 6 třídami se objevila velmi specifická malá třída (8 %) s přibližně průměrnou růstovou trajektorií a násobně větším reziduálním rozptylem, než mají ostatní třídy. Mohla by tohle být třída reprezentující jedince s vysoce fluktuující hodnotou autonomie? Z hlediska teorie by bylo výhodné moci takovou subpopulaci identifikovat. Ovšem to, že se třída příliš

neliší od ostatních svými růstovými parametry, komplikuje přesnou klasifikaci. I když na celkovém ukazateli separace tříd - entropii - se to podepsalo jen minimálně, klasifikační pravděpodobnosti pro tuto třídu jsou skokově nižší, než u ostatních tříd. U účastníků, kteří byli klasifikováni do této třídy, je jen 48% pravděpodobnost, že do této třídy skutečně patří. Je nezanedbatelná pravděpodobnost, že ve skutečnosti pocházejí ze sousedních větších tříd č. 5 a č. 6 (18 % resp. 27%). Lze tedy říci, že model se šesti třídami podporuje teoretizování o existenci subpopulace s vysoce volatlní prožívanou autonomií, ale nepodává o ní přesvědčivé důkazy. Tento výsledek však může inspirovat k neobvyklé specifikaci mixture modelu, kdy rezidua všech tříd vyjma jedné budou stejná, tedy k modelu, který v závěru pojmenovávám jako model hybridními třídami. Takový model by se snažil identifikovat třídy lišící se svými růstovými parametry a jednu třídu lišící se reziduálním rozptylem.

Stejně jako u lineárních modelů výše platí, že ani model se šesti třídami s téměř trojnásobkem volných parametrů nevykazuje stejně dobrou shodu s daty jako výchozí model latentních růstových křivek. Latentní růstové třídy tak můžeme považovat za alternativní, nikoli však lepší zachycení heterogenity individuálních růstových křivek.

Obrázek IV.10. Pozorované a odhadnuté průměry v LCGA lin. modelu se 6 třídami



LCGA KVADRATICKÉ MODELY S AUTOKORELAČNÍ REZIDUÁLNÍ STRUKTUROU

Posledním modelem růstu prožívané autonomie, na který aplikuji analýzu latentních růstových tříd, je kvadratický LGC model s autokorelacemi reziduí prvního řádu. Z výchozích modelů růstu prožívané autonomie měl tento model nejlepší shodu s daty při nízkém počtu parametrů (rozptyly i autokorelace byly modelovány jako konstantní).

U tohoto modelu jsem použil odlišný analytický postup, kdy jsem pro každý počet latentních tříd odhadoval dva i více modelů, které se lišily ve specifikaci odlišností reziduálních struktur napříč latentními třídami.

Modely, které mají v tabulce IV.11 ve svém názvu na konci „a“, jsou modely, v nichž jsou parametry reziduální struktury napříč latentními třídami stejné, homoskedastické. To znamená, že pro každou třídu se odhadují 4 parametry navíc - velikost třídy a tři průměry latentních růstových parametrů.

Modely, které mají ve svém názvu na konci „b“, jsou modely, v nichž se parametry reziduální struktury mohou napříč třídami lišit, jsou tedy heteroskedastické. To znamená ještě další dva parametry pro každou třídu navíc – rezidua (v čase konstantní) a reziduální autokovariance 1. řádu (také v čase konstantní). Celkem je to tedy 6 parametrů pro každou třídu navíc.

Modely s „c“ na konci jsou kombinací předchozích dvou, kdy jedné latentní třídě je umožněno mít své vlastní reziduální parametry, ale ostatní třídy již mají hodnoty reziduí a reziduálních kovariancí stejné. Jde tedy o specifikaci založenou na předpokladu, že většina tříd se bude lišit růstovou trajektorií, vyjma jedné, která nebude specifická svou růstovou trajektorií, ale reziduálními rozptyly – hypoteticky vyššími než v ostatních třídách. Z toho plyne, že model se dvěma třídami se od modelu s jednou třídou liší o 6 parametrů, další navyšování počtu tříd již znamená jen 4 parametry na další třídu. Model je tak úspornější. Ke specifikaci těchto modelů mě vedlo to, že rezidua lišící se mezi třídami jsou z hlediska odhadu velmi náročné a zvyšují množství konvergenčních problémů, a přitom se odhady těchto parametrů napříč třídami liší jen málo. Také, mimo třídu s vysoce variabilním vývojem prožívané autonomie, nemám dostatečné teoretické důvody pro hypotetizování rozdílných reziduí.

Tabulka IV.11. Ukazatele shody modelů s daty pro LCGA modely vycházející z M2cor

M2cor	par	LL	AIC	BIC	SSABIC	Entropie	LMR	LMR p	BLRT (Δ par)	BLRT p
LGC	11	-14537,3	29162	29217	29182					
C2a	9	-14747,1	29512	29557	29528	0,73	221,0	<0,0001	228,9 (4)	<0,0001
C2b	11	-14737,6	29497	29552	29517	0,73	221,2	0,003	226,5 (6)	<0,0001
C3a	13	-14632,6	29291	29356	29314	0,71	94,3	0,013	96,6 (4)	<0,0001
C3b	17	-14624,3	29283	29367	29313	0,70	92,6	0,006	95,9 (4)	<0,0001
C4b	23	-14576,0	29198	29312	29239	0,69	88,2	0,030	91,4 (4)	<0,0001
C4c	19	-14578,4	29195	29289	29229	0,69	32,9	0,065	34,1 (4)	<0,0001
C5c	23	-14532,7	29111	29225	29152	0,65	24,2	0,177	25,1 (4)	<0,0001
C6c	27	-14515,7	29085	29219	29133	0,68	18,7	0,583	19,4 (4)	<0,0001
C7c	31	-14503,1	29068	29222	29124	0,66	94,3	0,013	96,6 (4)	<0,0001

Likelihood testy – LMR a BLRT – vždy srovnávají daný model se stejným typem modelu, který má o jednu latentní třídu méně. Z hlediska modelů, jako jsou „c“, kde jsou odlišnosti mezi třídami stanoveny různě, je důležité si připomenout, že v Mplus likelihood testy při odhadu modelu s k-1 třídami „škrtají“ první specifikovanou třídu. Odlišně specifikovaná třída by tak neměla být tou první; já jsem si ji zvykl specifikovat jako poslední.

Z uvedených modelů vychází plauzibilně právě modely kombinující jednu třídu s odlišnými rezidui s více třídami se shodnými rezidui – hybridně specifikované latentní třídy. Až do modelu s celkem sedmi třídami se zlepšuje shoda modelu s daty, a přitom se specifická třída s odlišnými rezidui příliš nemění. Tato specifická třída (v tabulkách níže má číslo 1) reprezentuje okolo 10 % populace. Její růstová křivka je spíše průměrná s počátkem mezi 22 a 25 body a jen mírným růstem, maximálně 1 bod za rok (to vše na kvadratické škále autonomie od 1 do 49 bodů). Růst je přibližně lineární, protože kvadratický člen je blízký nule. Co je na

třídě specifické, je její reziduální rozptyl, který je násobkem reziduálních rozptylů v ostatních třídách. Hodnota rozptylu mezi 80 a 95 znamená přibližně směrodatnou odchylku reziduí 9-10. Třída tak reprezentuje jednotlivce, jejichž hodnoty autonomie mají sklon kolísat téměř v celém rozsahu škály. Ostatní třídy mají reziduální rozptyl zhruba čtyřikrát menší.

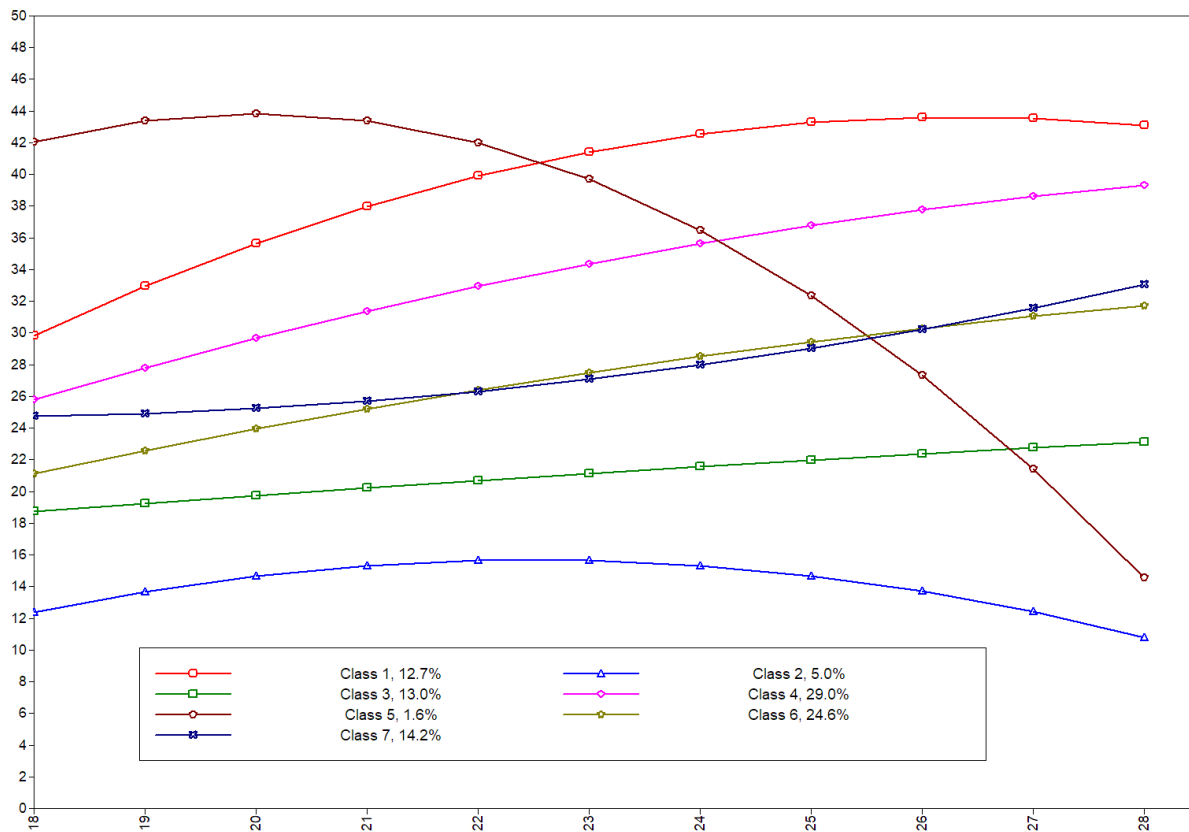
Tabulka IV.12. Parametry LCGA modelů vycházejících z M2cor

Model		C2a	C2b	C3a	C3b	C4b	C4c	C5c	C6c	C7c
<i>Relativní četnost</i>	C#1	0,46	0,50	0,32	0,31	0,08	0,08	0,12	0,13	0,14
	C#2	0,54	0,49	0,20	0,23	0,19	0,19	0,34	0,02	0,01
	C#3			0,48	0,46	0,31	0,30	0,12	0,32	0,12
	C#4					0,42	0,43	0,15	0,27	0,32
	C#5							0,28	0,11	0,14
	C#6								0,15	0,26
	C#7									0,05
<i>Růstové parametry</i>										
C#1	I	19,3	28,5	30,4	30,6	21,9	22,2	24,32	24,72	24,75
	S	0,96	1,92	2,06	2,03	0,86	0,95	0,63	0,03	0,10
	Q	-0,024	-0,087	-0,101	-0,099	0,002	-0,014	0,008	0,077	0,074
C#2	I	28,1	19,6	17,8	18,0	17,2	17,0	25,26	41,72	42,06
	S	1,90	1,00	0,13	0,24	0,62	0,66	2,12	2,50	1,80
	Q	-0,084	-0,024	0,029	0,022	-0,025	-0,029	-0,085	-0,607	-0,455
C#3	I			22,4	22,6	30,6	30,8	15,99	25,08	29,83
	S			1,48	1,54	2,00	1,98	0,45	2,06	3,31
	Q			-0,036	-0,040	-0,095	-0,093	-0,024	-0,075	-0,198
C#4	I					22,9	22,9	33,62	20,21	25,8
	S					1,42	1,44	1,84	1,34	2,07
	Q					-0,030	-0,035	-0,083	-0,042	-0,072
C#5	I							20,44	16,28	18,72
	S							1,27	0,20	0,52
	Q							-0,037	0,012	-0,008
C#6	I								29,99	21,12
	S								3,00	1,50
	Q								-0,169	-0,044
C#7	I									12,36
	S									1,47
	Q									-0,163
Reziduální rozptyly		39	34	32	29	94	95	88	83	80
		...	43	...	37	28	27	23	22	21
					31	28
						26				

Poznámka. Červeně zvýrazněné jsou problematické parametry.

Zajímavá je u těchto modelů jen velmi pomalu klesající hodnota entropie, která i při sedmi skupinách jen mírně klesá pod hodnotu 0,7. Třídy je tak možné považovat za daných okolností za poměrně dobře separované. Věcně třídy opět spíše pouze diskretizují normální rozložení růstových parametrů – třídy nad 10 % jsou přibližně paralelní. Smyslem tohoto řešení je tak spíše vyfiltrování části populace s rovnoměrným vývojem od části populace s vysoce fluktuujícím průběhem prožívané autonomie.

Obrázek IV.11. Vývojové trajektorie implikované LCGA kvadratickým modelem se sedmi třídami.



Poznámka. Pořadí tříd je zde jiné než v tabulkách. Specifická třída zde má číslo 7. Třídy jsou rozeznatelné nejnásne podle počáteční hodnoty – průsečíku.

Růstové mixture modely autonomie

LCGA modely popsané výše nebyly při popisování heterogenity růstových křivek příliš úspěšné. Většina modelů hůře či lépe popisovala spojité rozložení růstových křivek prostřednictvím tříd s víceméně paralelními růstovými křivkami lišícími se především počáteční hodnotou. Většina z těchto modelů nepopisovala data lépe než výchozí LGC modely, které růstové křivky popisují jako normální rozložení růstových parametrů. Z toho lze usuzovat, že rozdíly mezi individuálními růstovými křivkami prožívané autonomie v populaci jsou spíše spojité než diskrétní. Proto se zdá být smysluplnější pokusit se popsat heterogenitu pomocí jednoho (LGC) nebo směsi více normálních rozložení latentních růstových koeficientů (GMM). Oproti LCGA modelům se tedy GMM modely liší tím, že růstové parametry latentních tříd mají nenulový rozptyl. Individuální odchylky od třídní průměrné trajektorie pak nejsou považovány za rezidua a lze je pak v návazných modelech asociovat s korelátou či prediktory.

Díky nenulovým rozptylům latentních růstových koeficientů se rozšiřuje paleta voleb, které je potřeba učinit při specifikaci GMM modelu – mohou se napříč latentními třídami (komponentami směsi) lišit, mohou být stejné, nebo lze dokonce specifikovat „hybridní“ model, v němž se jedna ze tříd v těchto rozptylech liší od ostatních, které je mají stejné.

Tím, že je uvnitř tříd zachyceno větší množství rozptylu růstových koeficientů, obvykle klesá počet tříd potřebných pro zachycení heterogenity individuálních růstových křivek. S tím se také oproti výše uvedeným analýzám proměňuje analytická strategie, oproti výše prezentovaným LCGA modelům. Protože je zde pro každý počet tříd tolik možností, jak specifikovat rozdíly (či jejich absenci) mezi třídami, a protože odhady jednotlivých modelů s počtem tříd skokově narůstají, zdá se být vhodnější nejprve explorovat různě specifikované modely se dvěma třídami, pak se třemi třídami a tak dál. Dalším argumentem pro tento postup je to, že na této úrovni komplexity modelů mírně ustupují do pozadí ukazatele shody modelu s daty a testy rozdílu shody modelu s daty mezi modely s různými počty tříd. S rostoucím počtem tříd roste shoda modelu s daty často i při identifikačních a konvergenčních obtížích způsobujících nepřijatelná řešení. Přípustnost řešení a teoretická interpretovatelnost se pak stávají primárními kritérii hodnocení modelů.

GMM LINEÁRNÍHO RŮSTOVÉHO MODELU AUTONOMIE S HOMOSKEDASTICKÝMI REZIDUI

Pro aplikaci modelu směsí jsem zvolil nejjednodušší model vývoje autonomie, a to ten lineární s homoskedastickými rezidui. V tabulkách níže jsou uvedeny ukazatele shody s daty pro modely s různě specifikovanými rozdíly mezi latentními třídami a s různým počtem latentních tříd.

Modely s koncovým písmenem „a“ jsou modely, v nichž se latentní třídy lišily pouze průměrnými hodnotami latentních růstových koeficientů (tj. tvarem růstové křivky). Ve všech latentních třídách tak měly latentní růstové koeficienty stejný rozptyl. I reziduální rozptyly byly shodné napříč třídami. S každou latentní třídou tak přibývají 3 volné parametry. Takový model je konzervativním rozšířením LCGA modelu; třídy mají nějakou vnitřní heterogenitu, ale všechny stejnou. To odpovídá situaci, kdy nějaký faktor rozliší subpopulace s různými vývojovými trajektoriemi, ale další procesy/proměnné, které způsobují individuální variabilitu růstových křivek, jsou napříč subpopulacemi shodné. Stejně tak jsou shodné i procesy, které způsobují krátkodobé či náhodné fluktuace modelované charakteristiky kolem individuální růstové křivky. Právě takové očekávání neměnnosti většiny procesů, které se na pozorovaných datech podílí, činí takovou specifikaci dle mého názoru konzervativní. Její výhodou je však bezesporu parsimonie.

Modely s koncovým písmenem „b“ umožňují latentním třídám, aby se lišily i v rozptylu latentních růstových koeficientů. S každou latentní třídou přibývají ještě 3 volné parametry navíc (2 rozptyly a 1 kovariance), celkem 6. Takováto specifikace modelu je realističtější. Předpokládáme zde, že procesy, které se podílely na původním rozlišení do subpopulací (latentních tříd) stále běží a v jejich důsledku se liší variabilita růstových křivek mezi těmito subpopulacemi.

Modely s koncovým písmenem „c“ umožňují latentním třídám, aby se lišily ještě navíc v reziduálních rozptylech. S každou latentní třídou přibývá ještě 1 volný parametr navíc, celkem 8. Tím můžeme dát prostor předpokladu, že ani krátkodobě působící faktory způsobující odchýlení od individuální růstové křivky nejsou napříč subpopulacemi shodné. To může být jak „pozitivní“ předpoklad o nějakém faktoru, který tuto fluktuaci způsobuje, ale

také „negativní“ předpoklad o nedostatcích měřicího nástroje, který má v důsledku nedostatečné invariance měření napříč subpopulacemi v nich má různou reliabilitu.

Modely s koncovým písmenem „d“ umožňují latentním třídám, aby se lišily jen reziduálních rozptylech a nikoli v rozptylech latentních růstových koeficientů. To je celkem 5 volných parametrů na třídu. V modelech s více než dvěma třídami jsem takto specifikoval „hybridní“ modely, kde jsou rozptyly latentních koeficientů a reziduální rozptyly shodné napříč všemi třídami vyjma jedné. Tou jednou třídou by byla třída se zvýšeným reziduálním rozptylem.

Konečně, model s koncovkou „e“ je poněkud zvláštní, protože se v něm mohou latentní třídy lišit jen a pouze v reziduálních rozptylech, tedy ani v průměrech ani v rozptylech latentních růstových koeficientů. Tento model, zde odhadovaný pouze se dvěma třídami, by vycházel z předpokladu, že v populaci jsou subpopulace, které se neliší svou průměrnou růstovou trajektorií, ani její variabilitou, ale pouze náhodnými fluktuacemi kolem individuální růstové trajektorie. Protože průměry latentních růstových koeficientů se v Mplus defaultně liší, je potřeba je explicitně zafixovat.

Tabulka IV.13. Ukazatele shody modelů s daty pro GMM modely lineárního růstového modelu s homoskedastickými rezidui

M1_0	C	par	LL	AIC	BIC	SSABIC	Entropie	LMR	LMR p	BLRT	BLRT p
LGC	1	6		29279	29309	29290					
LGCA C6	6	18	-14696	29430	29519	29462	0,70	46,4	0,263	48,7 (3)	<0,0001
M1_G2a	2	9	-14623	29265	29309	29281	0,98			20,5 (3)	0,002
M1_G2b	2	12	-14617	29260	29319	29281	0,87	30,8	0,036	31,5 (6)	<0,0001
M1_G2c	2	13	-14546	29118	29183	29141	0,46			174,9 (7)	<0,0001
M1_G2d	2	10	-14547	29116	29165	29134	0,47	165,5	<0,0001	171,5 (4)	<0,0001
M1_G2e	2	8	-14551	29119	29159	29133	0,48	153,3	<0,0001	164,3 (2)	<0,0001
M1_G3a	3	12	-14616	29257	29317	29279	0,99			13,1 (3)	0,010
M1_G3b	3	18	-14608	29253	29343	29285	0,54			31,3 (6)	<0,0001
M1_G3c	3	20	-14536	29113	29212	29148	0,42	19,2	0,1778	19,6 (7)	0,052
M1_G3d*	3	16	-14539	29111	29190	29140	0,41	164,2	<0,0001	167,6 (7)	<0,0001
M1_G4d*	4	19	-14536	29112	29206	29146	0,41	156,5	<0,0001	159,7 (7)	<0,0001
M1_G5d*	5	22	-14535	29114	29223	29153	0,41				

Poznámka. Na prvních dvou řádcích pro srovnání výchozí LGC model a nejlepší LCGA model se 6 latentními třídami. LMR testy byly spočítány jen podle skutečné potřeby. * hybridní modely s pouze jednou třídou lišící se svými rozptyly.

Z hlediska porozumění heterogenitě růstových křivek prožívané autonomie nejsou modely „a“, v nichž se latentní třídy liší pouze průměrnými trajektoriemi, příliš přínosné. Model se dvěma i třemi třídami vlastně vyčleňují vedle majority pouze jednu nebo dvě mikroskopické latentní třídy, jakési outliery. I kvůli tomu není shoda těchto modelů s daty o nic lepší, než je původní LGG model.

Tabulka IV.14. Parametry GMM modelů lineárního růstového modelu s homoskedastickými rezidui

Model		G2a	G2b	G2c	G2d	G2e	G3a	G3b	G3c	G3d	G3e	G4d	G4e	G5d
<i>Relativní četnost</i>	C1	0,99	0,99	0,85	0,86	0,86	0,01	0,01	0,12	0,17	0,16	0,17	0,18	0,19
	C2	0,01	0,01	0,15	0,14	0,14	0,99	0,51	0,24	0,24	0,35	0,34	0,34	0,25
	C3						0,00	0,47	0,64	0,59	0,48	0,12	0,32	0,04
	C4											0,37	0,17	0,31
	C5													0,21
<i>Průměry růstových parametrů</i>														
C1	I	24,93	24,78	25,04	25,00	24,66	-15,89	19,26	24,22	23,42	23,68	23,68	23,68	23,43
	S	1,00	1,04	1,08	1,09	1,05	7,17	1,40	0,92	0,96	0,93	0,92	0,93	1,00
C2	I	-16,14	19,97	23,45	23,49	24,66	24,87	21,09	31,28	22,43	20,75	29,73	21,97	20,84
	S	7,19	1,37	0,94	0,90	1,05	1,01	0,86	1,12	0,18	0,73	1,39	0,47	0,67
C3	I						94,75	29,09	22,11	26,36	28,32	20,98	26,67	31,93
	S						-11,15	1,20	1,04	1,49	1,33	-0,11	1,15	-1,96
C4	I											21,99	28,21	21,78
	S											1,27	2,10	2,46
C5	I													31,64
	S													0,43
<i>Rozptyly růstových parametrů</i>														
C1	I	67,5	62,1	69,9	70,8	70,9	66,1	1098,0	83,4	76,9	76,9	76,1	72,9	80,7
	S	1,8	1,6	1,7	1,8	1,8	1,7	27,6	1,7	1,8	1,8	1,8	1,8	1,9
C2	I	67,5	1084,3	76,4	70,8	70,9	66,1	43,5	51,3	66,2	52,8	54,5	54,9	43,2
	S	1,8	27,6	2,0	1,8	1,8	1,7	1,9	1,4	1,4	2,0	1,4	1,7	0,1
C3	I						66,1	49,5	51,1	66,2	58,2	54,5	84,3	43,2
	S						1,7	1,3	1,9	1,4	1,4	1,4	1,4	0,1
C4	I											54,5	47,9	43,2
	S											1,4	0,6	0,1
C5	I													43,2
	S													0,1
<i>Kovariance růstových parametrů</i>														
C1		-6,0	-5,1	-5,9	-6,4	-6,3	-5,6	-167,7	-9,2	-8,7	-8,8	-8,6	-8,1	-9,2
C2		-6,0	-166,4	-8,7	-6,4	-6,3	-5,6	-5,7	-6,6	-6,9	-7,3	-7,1	-6,6	-1,5
C3							-5,6	-5,8	-5,7	-6,9	-6,7	-7,1	-10,2	-1,5
C4												-7,1	-4,9	-1,5
C5														-1,5
<i>Reziduální rozptyly</i>														
C1		19,6	19,4	12,4	12,6	12,7	19,5	19,4	51,7	46,4	47,1	47,7	47,4	45,6
C2		19,6	19,4	46,1	47,2	48,2	19,5	19,4	10,2	12,4	12,5	12,6	12,6	12,3
C3							19,5	19,4	14,7	12,4	12,5	12,6	12,6	12,3
C4												12,6	12,6	12,3
C5														12,3

Poznámka. Červeně zvýrazněné jsou problematické parametry.

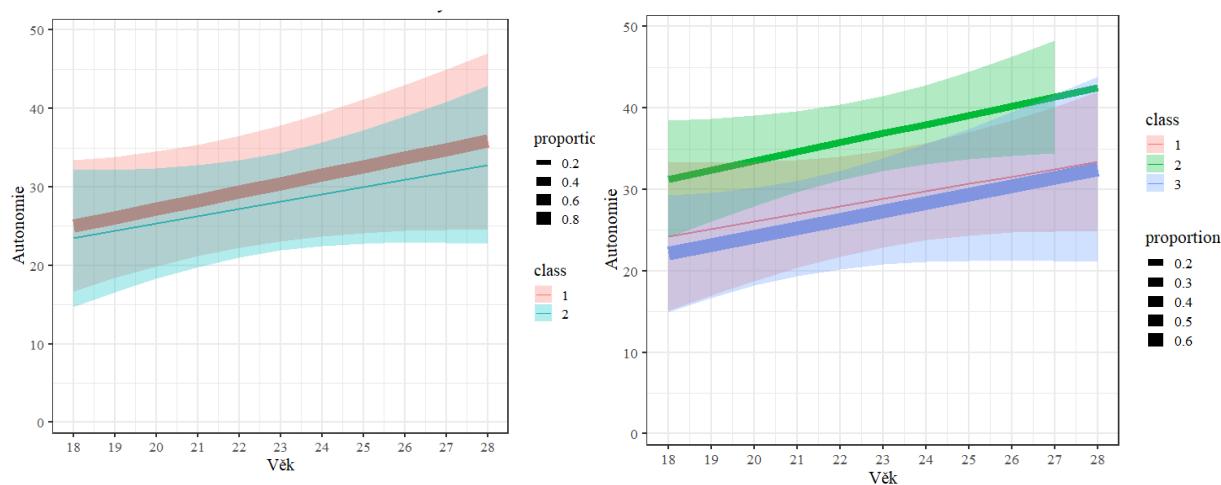
Modely „b“ se třídami lišícími se pouze latentními růstovými koeficienty, jejich průměry, rozptyly a kovariancí, jsou na tom podobně. Model se dvěma latentními třídami je prakticky identický variantě modelu „a“, přičemž velikost malé latentní třídy je tak malá, že je odhad

rozptylů latentních koeficientů v této třídě zatížen obrovskou chybou. V případě modelu se třemi třídami se totéž přihodilo s jednou z mikroskopických tříd, přičemž ta druhá byla odhadnuta jako totožná s majoritou (respektive vznikla přibližným rozpůlením majority). Další navýšení počtu parametrů nevedlo k lepší shodě modelu s daty, a tak informační ukazatele penalizující za komplexitu modelu vychází pro tento model ještě horší než pro modely „a“. Zdá se tedy, že bez nějakého umožnění rozdílů v reziduálních strukturách mezi latentními třídami nelze heterogenitu růstových křivek popsat lépe, než to činí prostý LGC model.

Modely „c“, které umožňují odlišnosti reziduí napříč třídami, již nabízí informativnější obrázek. Třídy mají plauzibilní velikosti a informační ukazatele i přes nárůst počtu parametrů vypovídají o lepší shodě těchto modelů s daty. Ovšem daní za umožnění odlišných reziduí je dramatický pokles entropie do hodnot kolem 0,4, ty umožňují další modelování, ale rozhodně ne klasifikaci, která by byla použitelná mimo model, bez informace o pravděpodobnosti členství v jednotlivých třídách. To je charakteristické i pro všechny další modely prezentované níže. Rozptyl reziduí a rozptyl růstových křivek jsou v případě autonomie patrně částečně závislé, a tak je kategorizování podle obou kritérií najednou nejisté a pro jeho úplnou realizaci by bylo potřeba větší množství latentních tříd, pro jejichž identifikaci však je patrně potřeba více dat, než máme k dispozici. Navíc je potřeba připomenout zatím nepotvrzený předpoklad o existenci latentních tříd. GMM není klasifikace dle kritéria, ale klasifikaci „dle potřeby“, potřeby dat.

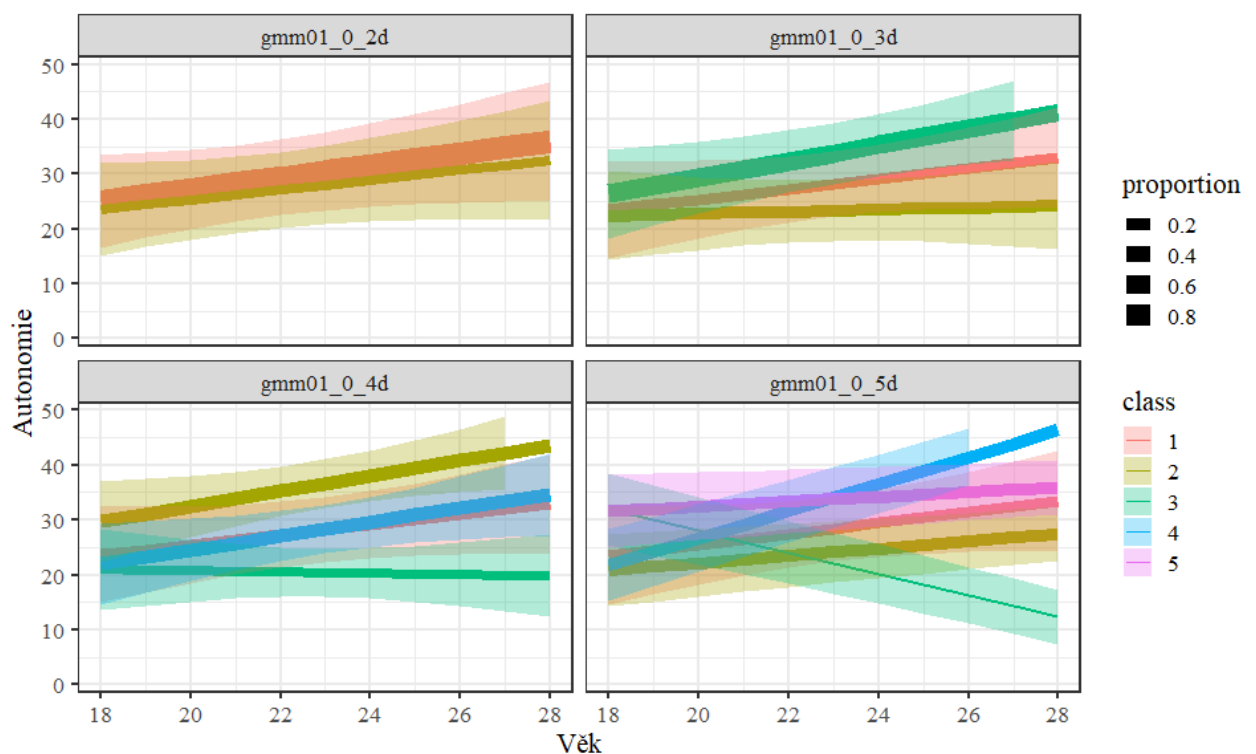
Z modelů „c“ se jeví jako zajímavější model se třemi třídami, kdy třídy 1 a 3 mají velmi podobnou střední vývojovou trajektorii i podobný rozptyl trajektorií uvnitř třídy, ale liší se reziduálním rozptylem, který je v poměrně menší první třídě několikanásobně vyšší než ve třídě 3. Třída 2 reprezentující asi čtvrtinu populace se pak odlišuje od třídy 3 především vyšší průměrnou trajektorií prožívané autonomie, rozptylové parametry mám podobné. Jsou zde tedy tři třídy – majorita se střední a rostoucí mírou autonomie (64 %), minorita s podobně rostoucí autonomií, ale skokově vyšší mírou fluktuace kolem individuálních růstových křivek (12 %) a konečně třída s vyšší a také rostoucí mírou autonomie (24 %) a podobnou mírou fluktuace, jako nalézáme u majority.

Obrázek IV.12. Latentní třídy v modelech 2c a 3c



Modely „c“ s vyšším počtem tříd již neposkytovaly přípustná řešení. Cesta dál vedla přes omezení možnosti rozptylových parametrů lišit se napříč třídami. Protože specifikace neumožňující žádné rozdíly v reziduálních rozptylech napříč třídami neskýtaly použitelná řešení, rozhodl jsem se omezit rozptyly latentních růstových koeficientů. Ovšem u reziduálních rozptylů se vynořila řada konvergenčních problémů, pokud se měly lišit napříč více než dvěma latentními třídami. Proto má model „d“ se třemi a více třídami „hybridní“ specifikaci, kdy se jedna ze tříd liší od zbývajících dvou, a to jak v reziduálních rozptylech, tak v rozptylech latentních růstových koeficientů. Tyto modely nabízejí analyticky zajímavá řešení, v nichž se jako první třída opakuje třída se střední počáteční mírou autonomie a středním tempem růstu přibližně o 1 bod za rok na kvadratické škále autonomie. Tato třída, odhadovaná jako stále přibližně stejně velká (neceľých 20 % populace), má přibližně čtyřnásobný reziduální rozptyl než ostatní třídy. Ten nabývá hodnoty necelých 50, což znamená směrodatnou odchylku o hodnotě přibližně 7. Jsou-li tedy rezidua normálně rozložená většina fluktuace kolem individuální růstové křivky je v pásmu 14 bodů nad a pod, což je velmi široké pásmo vzhledem k celkovému rozpětí škály kvadratické autonomie od 1 do 49. Ostatní třídy mají reziduální rozptyl mezi 12 a 13, což znamená fluktuace v pásmu přibližně 7 bodů nad a pod individuální růstovou křivkou na kvadratické škále autonomie. Další latentní třídy pokrývají rozpětí pozorovaných individuálních růstových křivek, jak je zobrazeno v grafech na Obrázku IV.13.

Obrázek IV.13. GMM modely lineárního růstu s „hybridními“ třídami



Volba mezi modely se 3, 4 nebo 5 třídami je zde pak obtížná. Informační kritéria preferují spíše model se třemi třídami, LMR a BLRT testy model se 4 třídami, přičemž pro model s 5 třídami BLRT test vycházel s chybami, takže jej nelze považovat za vypovídající. Ani hodnota entropie zde nepomůže, protože má u všech modelů nízkou hodnotu 0,4. Nejistotu spojenou

s modelem s 5 třídami odráží i to, že se oproti modelům s méně třídami podoba tříd výrazně proměnila. V této situaci bych preferoval model se 4 třídami, který vedle specifické třídy s vysokým reziduální rozptylem nalézá tři třídy, z nichž dvě začínají na podobné hodnotě autonomie (21-22) a liší se svým růstem. Jedna mírně roste a druhá stagnuje. S ohledem na rozptyl latentních směrnic (1,4) je zřejmé, že se třídy překrývají, což je vidět i na obrázku IV.13 a podílí se na nízké hodnotě entropie. Zbývající třída má o něco vyšší počáteční hodnotu autonomie a také mírný růst. Se sousední třídou se také mírně překrývá.

Prizmatem výše popsaných modelů směsi lineárních růstových křivek autonomie lze uvažovat tak, že rozložení individuálních růstových křivek je spojitě a podobné normálnímu natolik, že jeho popis prostřednictvím pouze jedné třídy (tedy LGC modelem) je do značné míry adekvátní. Je-li zde co kategorizovat, nejsou to ani tak růstové křivky samotné, ale reziduální rozptyl kolem nich. Pro to svědčí i fakt, že triviální model se dvěma třídami, které se mohou lišit pouze reziduálním rozptylem, má z všech modelů prezentovaných v této sekci nejlepší hodnoty infomačních ukazatelů penalizujících za komplexitu modelu. Je to tedy ten nejlepší jednoduchý model. Modely s více třídami pak dále kategorizují tu část populace, v níž je kolísání hodnot kolem individuálních růstových křivek nižší, nikoli však zanedbatelné. V modelu se čtyřmi třídami (1+3) do dvou rostoucích tříd – střední a vyšší – a jedné stagnující. To jsou závěry, které se podobají zjištěním z použití LCGA modelů. I u těch bylo pro shodu modelu s daty důležité umožnit alespoň jedné třídě odlišné reziduální rozptily.

Z uvedeného je zřejmé, že rezidua hrají v popisu heterogenity vývojových křivek prožívané autonomie zásadní roli. Přitom ve všech výše popsaných modelech se na jejich velikosti podílí celá řada faktorů. Tím prvním je vůbec tvar/forma modelovaných růstových křivek. Je zřejmé, že oproti přímkám mají křivky vyšší schopnost popsat individuální růst, což implikuje nižší rezidua. Druhým jsou faktory a procesy, které se podílejí na krátkodobých změnách, či fluktuacích modelované proměnné. Může jít jak o vlivy prostředí, tak například o osobnostní či osobní proměnné, které spoluurčují dynamiku prožívání. Ty lze v případě dostupnosti dat zařadit do LGC modelu. Ovšem v případě GMM modelů je zařazení v čase se měnících kovariátů poměrně náročné. Konečně třetím faktorem je reliabilita měření, jejíž nedostatek může vytvořit zanedbatelné množství náhodného šumu kolem individuální růstové křivky. Zatímco efekt tvaru křivky je zřejmý z prezentovaných analýz a srovnání lineárního a kvadratického modelu vývoje autonomie, popř. různých forem vývoje rizikového chování o kapitole výše, v následující kapitole bych se rád pokusil zohlednit v modelu poslední faktor ovlivňující velikost reziduí, a to reliabilitu měření. Lze toho dosáhnout tím, že namísto manifestních skóru autonomie budu modelovat vývoj latentních skóru autonomie.

Růstový model druhého řádu – Curve-of-factors model

Všechny předchozí analýzy vývoje prožívané autonomie v rané dospělosti modelovaly vývoj manifestní proměnné – skóru autonomie vytvořeného zprůměrováním odpovědí na 7 položek škály autonomie (a následným umocněním na druhou pro korekci zešíkmení). Takový postup umožňuje komplexnější specifikaci modelu latentních růstových křivek, zvláště v případě longitudinálního designu s plánovaně chybějícími daty. Tím je v modelech také více prostoru pro komplexnější specifikaci tříd v mixture modelech. Dochází však k zanedbání informace o modelu měření autonomie. Reziduální rozptily v jednotlivých věcích reprezentují rozptyl nevysvětlený individuální růstovou křivkou jak z toho důvodu, že individuální křivka přesně nevystihuje individuální vývoj (není to přímka, parabola...), ale také z náhodných důvodů, kvůli chybám měření. Protože model s manifestními proměnnými nepracuje s reliabilitou

měření, není možné rozlišit tyto dvě složky reziduálního rozptylu. A nejistota se netýká jen reliability. Pokud měřítko autonomie není striktně unidimenzionální, je možné, že některé části pozorovaného vývoje nejsou vývojem autonomie, ale třeba vývojem specifických položkových faktorů, nebo jiných nežádoucích systematických vlivů. Proto je žádoucí mít model měření součástí růstového modelu a modelovat tak vývoj latentních skóre autonomie.

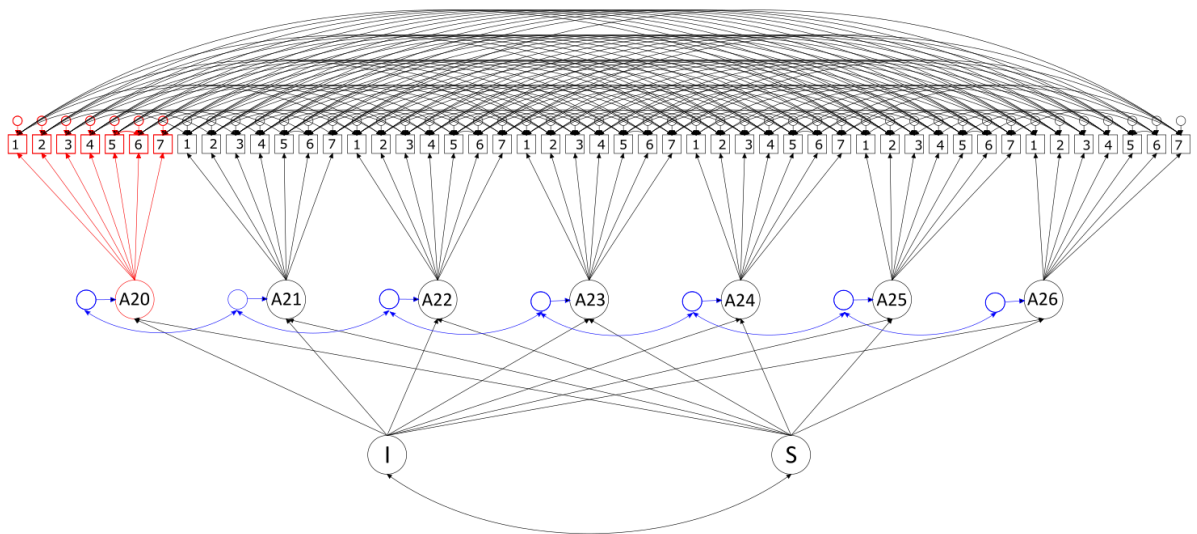
Zakomponování modelu měření do LGC modelu je ideově velmi jednoduchou myšlenkou s řadou technických detailů komplikujících specifikaci modelu a odhad jeho parametrů. Poměrně přímočarý je tzv. Curve-of-Factors model (CUFFS dle McArdle, 1988, CFM dle Wickrama et al., 2016), v němž je růstová křivka specifikována stejně jako v běžném LGC modelu s tím rozdílem, že latentní průsečík a směrnice determinují latentní skóre prožívané autonomie namísto manifestních (viz obrázek IV.14). V tomto modelu je autonomie v každém věku reprezentována modelem měření, v němž faktor autonomie sytí odpovědi na všech 7 položek (na obrázku IV.14 červeně). Aby bylo možné zaznamenat vývoj hodnot latentní proměnné autonomie, je nutné, aby byl model měření ve všech věcích stejný, alespoň na úrovni silné invariance. To je nejen požadavek modelu, ale především výhoda – zatímco validita manifestního skóre se může s věkem proměňovat (pokud by škála nebyla časově invariantní), latentní skóre budou z hlediska validity reprezentovat stále totéž a případný nedostatek invariance by se projevil zhoršením shody modelu s daty. Náboje a průsečíky položek (průsečíky v obrázku nejsou) jsou tedy v tomto vývojovém modelu fixovány na stejné hodnoty ve všech věcích. Položkové reziduální rozptyly mohou být napříč věky ponechány volné. V modelu měření autonomie je oproti čistě unidimenzionálnímu modelu navíc kovariance mezi rezidui sousedících položek 5 a 6. Ta není vysoká, v korelační matici nabývá v různém věku hodnot mezi 0,06 a 0,17. I ta může být ponechána volná napříč věky. I ve zohlednění této lokální závislosti položek se liší latentní skór autonomie od manifestního součtového skóre. Zatímco vývoj manifestního skóre může být ovlivněn vývojem toho, co mají položky 5 a 6 společného nad rámec toho, že jsou syceny autonomií, latentní skór reprezentuje jen to, co má všech sedm položek společného.

I když reziduální rozptyly položek nejsou v růstovém modelu primárním předmětem zájmu, je nutné v modelu specifikovat jejich korelace v čase. Obvykle je na místě předpokládat, že reziduální rozptyl položky je ze značné části unicitá (specifický faktor) s nějakou stabilitou v čase. Nejčastějším zohledněním této stability jedinečných faktorů je umožnění korelací mezi reziduálními rozptyly téže položky napříč opakovanými měřeními, jak je tomu i v modelu na obrázku IV.14. Při větším množství měření lze alternativně parametricky úsporněji specifikovat „položkový faktor“ jako latentní proměnnou sytící všechna opakovaná měření dané položky. Nezohlednění stability specifických faktorů vede ke zhoršení shody modelu s daty, což pak komplikuje další modelovací kroky, a může vést i ke zkreslení faktorových nábojů (McArdle & Nesselroade, 2014). Tyto longitudinální kovariance se opět mohou v čase lišit, byť v případě omezené velikosti vzorku lze uvažovat o omezenější reziduální struktuře.

V případě modelu vývoje prožívané autonomie jsem tedy specifikoval model se sedmi měřeními od 20 do 26 let (což je maximum umožněné zrychleným longitudinálním designem běžícím 4 roky), v němž jsou položková rezidua volně korelovaná napříč věky. Protože je cílem modelovat vývoj úrovně prožívané autonomie a nikoli pouze kovarianční vztahy, je nutné, aby byla dobře identifikovaná i struktura průměrů latentních proměnných. Toho je dosaženo tím, že průsečík markerové položky (jejímž nábojem fixovaným na 1 přidělujeme latentní proměnné metriku) fixujeme na hodnotu 0 v každém opakovaném měření/věku. Tím vlastně faktor přebírá průměr marker proměnné. Výsledkem je vlastně model se sedmi korelovanými měřeními autonomie od 20 do 26 let, kdy průměry latentních proměnných

reprezentujících autonomii v jednotlivých věcích popisujíc vývoj průměru prožívané autonomie.

Obrázek IV.14. Curve-of-factors model prožívané autonomie od 20 do 26 let.



Posledním krokem je specifikace modelu latentních růstových křivek. Ta je analogická situaci, kdy specifikujeme model růstových křivek na manifestních proměnných. V případě lineárního růstového modelu tedy specifikujeme latentní průsečík predikující všechna latentní měření autonomie s regresním koeficientem fixovaným na hodnotu 1 a latentní směrnici predikující všechna latentní měření autonomie s regresními koeficienty pevně fixovanými na hodnoty reflektující tok času, tedy v roli časových bází. Latentní průsečík (I na obrázku IV.14) je tedy proměnnou, která reprezentuje individuální počáteční úroveň autonomie a latentní směrnice (S) reprezentuje individuální tempo růstu za časovou jednotku danou tím, jak jsou specifikovány časové báze. Fixované regresní koeficienty latentního průměru a směrnice jim stanovují metriku, a tak zbývá jim ještě identifikovat průměr. Toho je dosaženo tím, že průměry latentních skóre autonomie jsou ve všech věcích fixovány na hodnotu 0. Tím se „přenes“ jejich průměrná hodnota do latentního průsečíku – pokud jsou časové báze specifikovány od 0, pak bude mít latentní průsečík průměr rovný průměru autonomie v prvním měření, tedy 20 letech.

Další prvky modelu latentních růstových křivek faktorů (curve-of-factors model, CFM) jsou také analogické klasickému LGC modelu. Latentní průsečík a směrnice jsou proměnné s nenulovým rozptylem, takže lze odhadovat i jejich kovarianci. Růstový model samozřejmě nevysvětluje individuální vývoj latentních skóre beze zbytku – ten reprezentují disturbance latentních skóre autonomie v jednotlivých letech. Ty jsou analogické reziduím v LGC modelu (na obrázku IV.14 modře) a je tedy na místě uvážlivě specifikovat i jejich vztahy. V modelu na obrázku IV.14 je specifikována velmi jednoduchá lag-1 autokorelační struktura, která odpovídá představě, že kromě lineárního vývoje má hodnota latentní autonomie nějakou setrvačnost, která ale nemá delší trvání než jeden rok.

PŘÍPRAVA DAT

Protože data z projektu Cesty do dospělosti obsahují více než jedno měření ročně, bylo stejně jako u manifestních skóre autonomie potřeba nalézt způsob agregace, který by

znamenal co nejmenší ztrátu informace. Zatímco manifestní skóry autonomie lze v rámci ročních intervalů zprůměrovat, u položkových dat, která jsou ve své podstatě ordinální kategorická, je tato operace problematictější. Položky jsou hodnoceny na sedmibodové škále, což znamená, že při absenci extrémního zešikmení či strmosti je lze bez významného zkreslení modelovat jako spojité manifestní proměnné (viz např. Maydeu-Olivares, Shi, & Rosseel, 2018). Průměrování však znamená potlačování náhodného rozptylu, vyhlazování individuální růstové křivky. Proto jsem se v případě položek, pro kontrast k předchozím analýzám a pro zachování možnosti pracovat s položkami jako s kategorickými ordinální proměnnými, rozhodl namísto průměrování odpovědí za roční interval náhodně vybrat jedno z dostupných vyplnění škály autonomie v daném roce. Celkově jsem udělal pět takových náhodných výběrů. Vždy se vybírala celá sada sedmi odpovědí ze stejného dotazníku i s případnými chybějícími odpověďmi. Všechny následující analýzy jsou provedeny na třetí sadě; zbývající byly uloženy pro případnou krosvalidaci (byť je mezi výběry značný překryv daný tím, že většina respondentů má v některém věku pouze jeden vyplněný dotazník – typicky na začátku či konci projektu).

LONGITUDINÁLNÍ INVARIANCE

Pro možnost modelování latentních růstových křivek z latentních proměnných je nutné, aby byl model měření proměnné, jejíž vývoj se snažíme modelovat, longitudinálně alespoň silně invariantní. To znamená, že náboje položek na latentní proměnné a jejich průsečíky by měly být konstantní napříč vlnami měření nebo věky (Y. Liu et al., 2017; Millsap, 2010). Tím je zajištěno to, že latentní proměnná reprezentuje měřený konstrukt stále stejně, nebo, jinými slovy, že změna hodnoty latentní proměnné nereprezentuje namísto individuálního vývoje posun ve způsobu měření. Výše popsany CFM model již longitudinální invarianci předpokládá; je však nutné ověřit, že tento předpoklad platí. Kdyby neplatil, jednak by to zhoršovalo shodu modelu s daty a komplikovalo rozhodování o rozdílech mezi složitějšími modely, a také by hrozilo riziko toho, že budeme modelovat vývoj více či méně odlišné proměnné. Odhad parametrů by totiž směřoval k nalezení takových nábojů a průsečíků položek, které by měly k naplnění invariance nejbližší. To může v extrému znamenat „vykousnutí“ velmi malé části konstruktů, která je z hlediska modelu měření v čase invariantní.

Tabulka IV.15. Popisné statistiky položek škály autonomie od 20 do 26 let (sada 3)

Věk	20			21			22			23		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD
A1	401	5,36	1,28	783	5,47	1,23	1015	5,58	1,17	875	5,76	1,12
A2	401	2,97	1,46	781	3,01	1,47	1013	2,91	1,49	877	2,79	1,45
A3	401	5,93	1,24	777	5,97	1,16	1013	5,94	1,12	874	6,04	1,06
A4	399	3,08	1,40	780	3,14	1,41	1014	3,01	1,39	874	2,86	1,40
A5	400	4,72	1,47	779	4,91	1,36	1013	4,98	1,36	871	5,17	1,32
A6	401	4,78	1,61	779	4,83	1,57	1012	4,88	1,58	876	5,03	1,61
A7	401	2,59	1,37	782	2,63	1,35	1014	2,48	1,25	877	2,28	1,21

Věk	24			25			26		
	N	M	SD	N	M	SD	N	M	SD
A1	704	5,82	1,11	477	5,96	1,11	263	5,96	1,04
A2	706	2,75	1,43	478	2,59	1,44	264	2,58	1,41
A3	703	6,08	1,05	474	6,10	1,10	264	6,14	1,01
A4	705	2,70	1,31	478	2,69	1,38	264	2,62	1,41
A5	703	5,18	1,31	479	5,37	1,28	265	5,38	1,24
A6	705	5,08	1,52	477	5,24	1,51	265	5,25	1,48
A7	706	2,30	1,28	479	2,16	1,21	265	2,26	1,30

Poznámka. Položky 2, 4 a 7 jsou opačně orientované. Všechny položky měly minimum 1 a maximum 7.

Vzhledem ke zrychlenému longitudinálnímu designu projektu Cesty do dospělosti není možné jednoduše odhadnout model pokrývající všechny věky najednou. Zejména v základním modelu měření, v němž jsou náboje i průsečíky volně odhadovány v každém věku, nemáme žádná data pro vztahy mezi prvky modelu (latentní autonomie, položky autonomie), které jsou od sebe v čase vzdálené více než 4 roky. I když parametry modelu implikují, že mezi autonomií například ve 20 letech a v 25 letech je nějaká korelace, v datech není nikdo, kdo by měl měření z 20 i 25 let, protože projekt tak dlouho neběžel. Algoritmus odhadu parametrů se s tím do jisté míry dokáže vyrovnat za cenu zhoršení ukazatelů shody modelu s daty, ale při postupném natahování v čase model přestane konvergovat. Pro stanovení longitudinální invariance jsem tedy modeloval zvlášť tři překrývající se období – od 19 do 22, od 22 do 25 a od 24 do 27. V 19 a 28 letech už bylo příliš málo dat. Tabulka IV.16 prezentuje shrnutí základního (konfigurálního), slabě a silně invariantního modelu v každém věkovém rozpětí. V každém věku byl model měření identický – jeden faktor, jehož metrika je dána fixováním náboje položky A1⁸ na 1 a jejího průsečíku na 0, reziduální kovariance mezi položkami 5 a 6. Longitudinálně bylo umožněno reziduím téže položky korelovat a korelace mezi měřeními autonomie měly autokorelační strukturu.

Všechny modely byly odhadovány s položkami jako spojitými proměnnými s použitím robustního full-information maximum likelihood estimátoru v Mplus 8 (MLR), který umožňuje chybějící hodnoty v datech. Ve všech třech obdobích lze model měření označit za silně longitudinálně invariantní. Fixování nábojů nezpůsobuje signifikantní zhoršení shody modelu s daty prizmatem korigovaného LRT testu dle (Bryant & Satorra, 2012). RMSEA se nezhoršuje, CFI klesá maximálně o dvě tisíciny, což je zcela zanedbatelná hodnota. SRMR se mění podobně minimálně a u posledního věkového rozpětí se dokonce mírně zlepšuje. Na velikost vzorku korigovaný BIC (SABIC) vždy klesne, tedy zlepšuje se. Následné fixování průsečíků sice vede k signifikantnímu zhoršení shody modelu s daty, ale jde o zhoršení naprosto minimální, jak ukazují všechny indexy shody. RMSEA se opět nezhoršuje, CFI klesá maximálně o 5 tisícín, SRMR roste maximálně o tisícínu a SABIC dále zřetelně klesá. Lze tedy říci, že náboje a průsečíky položek se v průběhu času uvnitř modelovaných čtyřletých intervalů významně nemění a jejich zafixování v růstovém modelu nemůže vést ke změně validity latentního skóru autonomie. Pohled na nestandardizované náboje a průsečíky položek ze tří silně invariantních modelů ukazuje, že předchozí tvrzení lze rozšířit na celé

⁸ Tato položka byla zvolena za marker, protože v předběžných analýzách vycházely její náboje i průsečíky v čase velmi stabilní.

sledované období od 19 do 27 let. Náboje se liší o pár setin, přičemž směrodatné chyby odhadu těchto parametrů se pohybují kolem 0,05

Tabulka IV.16. Shrnutí tří sekvenčních analýz longitudinální invariance měření prožívané autonomie

Věk	19-22			22-25			24-27		
	Konf.	Slabá	Silná	Konf.	Slabá	Silná	Konf.	Slabá	Silná
Chi2	491	508	551	570	587	628	556	567	587
df	304	322	340	304	322	340	304	322	340
sf	1,1	1,11	1,1	1,12	1,12	1,12	1,05	1,06	1,05
<i>p</i> (LRT)		0,774	0,000		0,650	0,006		0,961	0,192
RMSEA	0,021	0,021	0,021	0,026	0,026	0,026	0,033	0,032	0,031
CFI	0,961	0,961	0,956	0,960	0,960	0,956	0,926	0,928	0,928
SRMR	0,052	0,056	0,056	0,044	0,047	0,048	0,115	0,107	0,108
SABIC	61474	61420	61393	68067	68014	67985	32221	32176	32131
Invariantní náboje									
A1			1,00			1,00			1,00
A2			-1,15			-1,18			-1,14
A3			0,91			0,86			0,85
A4			-1,07			-1,08			-1,02
A5			0,72			0,83			0,84
A6			0,75			0,87			0,84
A7			-1,02			-1,06			-1,03
Invariantní průsečíky									
A1			0,00			0,00			0,00
A2			9,24			9,58			9,40
A3			0,94			1,10			1,09
A4			8,94			9,06			8,71
A5			0,93			0,34			0,34
A6			0,66			-0,02			0,21
A7			8,10			8,40			8,27
N ve věku									
	19	254		22	1013		24	720	
	20	499		23	904		25	483	
	21	893		24	720		26	277	
	22	1013		25	483		27	60	

Za zmínku stojí, že uvolnění autokorelační struktury latentních proměnných autonomie, což zvýší počet volných parametrů o tři, vede ke zcela minimálnímu poklesu chí-kvadrátu (o méně než 2). Minimálně uvnitř těchto čtyřletých úseků se zdá být vývoj autonomie velmi stabilní. Hledání rozdílů na hranici chyb měření by se mohlo zaměřit na položky 5 a 6, které mají mírně nižší náboje v prvním věkovém rozpětí, ale v kontextu plánovaného růstového modelu jde o rozdíl zanedbatelný. Průsečíky se liší o desetiny, ale i jejich standardní chyba odhadu se pohybuje kolem 0,3. I zde lze uvažovat o odlišnosti ranných průsečíků položek 5 a 6.

RŮSTOVÝ MODEL LATENTNÍ AUTONOMIE (CFM)

Růstový model má základní specifikaci, jak je uvedeno výše a na obrázku IV.14. Jde o vývoj od 20 do 26 let. I když fixování modelu měření napříč lety a fixování kovariancí v čase vzdálených reziduí umožňuje rozšířit časové rozpětí modelu vysoko nad délku sběru dat, rozpětí od 20 do 26 let je maximem, které umožňuje konvergenci a stabilní odhad parametrů i pro následné mixture modely. Růstová část modelu je specifikována jako lineární, tedy s jedním náhodným latentním průsečíkem (I) a jednou náhodnou latentní směrnici (S). Volba časových bází pro směrnici pak určuje, zda bude modelovaný růst vskutku lineární (báze od 0 do 6), odhadovaný (báze 0 na začátku, 6 na konci a zbytek odhadovaný), nebo třeba logistický (báze vypočítané na základě odhadnutých dvou parametrů logistického růstu). Pro kvadratický růst by již bylo potřeba do modelu přidat další náhodnou latentní proměnnou (Q).

Tabulka IV.17 shrnuje základní parametry čtyř odhadnutých CFM modelů – prostého lineárního růstu (M1), latent-base modelu, který volně odhaduje latentní báze pro věky 21-25 (M2), kvadratického růstu (M3) a logistického růstu. Všechny modely mají v čase konstantní disturbance s lag-1 autokorelační strukturou. Všechny uvedené modely mají velmi podobnou shodu s daty. Tu lze považovat za uspokojivou, protože mírně zvýšené CFI a SRMR lze připsat na vrub tomu, že v pozorované kovarianční matici jsou neodhadnutelné kovariance mezi věky, které jsou od sebe vzdálené více než o 5 let, přičemž model implikuje nenulové kovariance i mezi těmito věky.

Ve srovnání s růstovým modelem na manifestních součtových skórech autonomie je zde o něco méně variability, heterogenity k vysvětlení. V jednoduchém lineárním modelu má střední trajektorie počátek na hodnotě 5,4 a roste o 0,1 za rok. Směrodatná odchylka průsečíků je 0,7 a směrnice 0,1 se střední korelací -0,4. Pásmo tvořené křivkami s počátečními hodnotami 2 směrodatné odchylky nad a pod průměrnou trajektorií pokrývá horní polovinu škály – od 4 do 7. I když model růstu faktorů pokrývá kratší období (model s manifestními skóry bylo možné odhadnout pro plné rozpětí 18-28), je zde vidět, že část rozptylu odpovědi respondentů, kterou klasický LGC model modeloval jako součást reliabilního rozptylu, se v modelu křivek faktorů přesunula do reziduální části modelu.

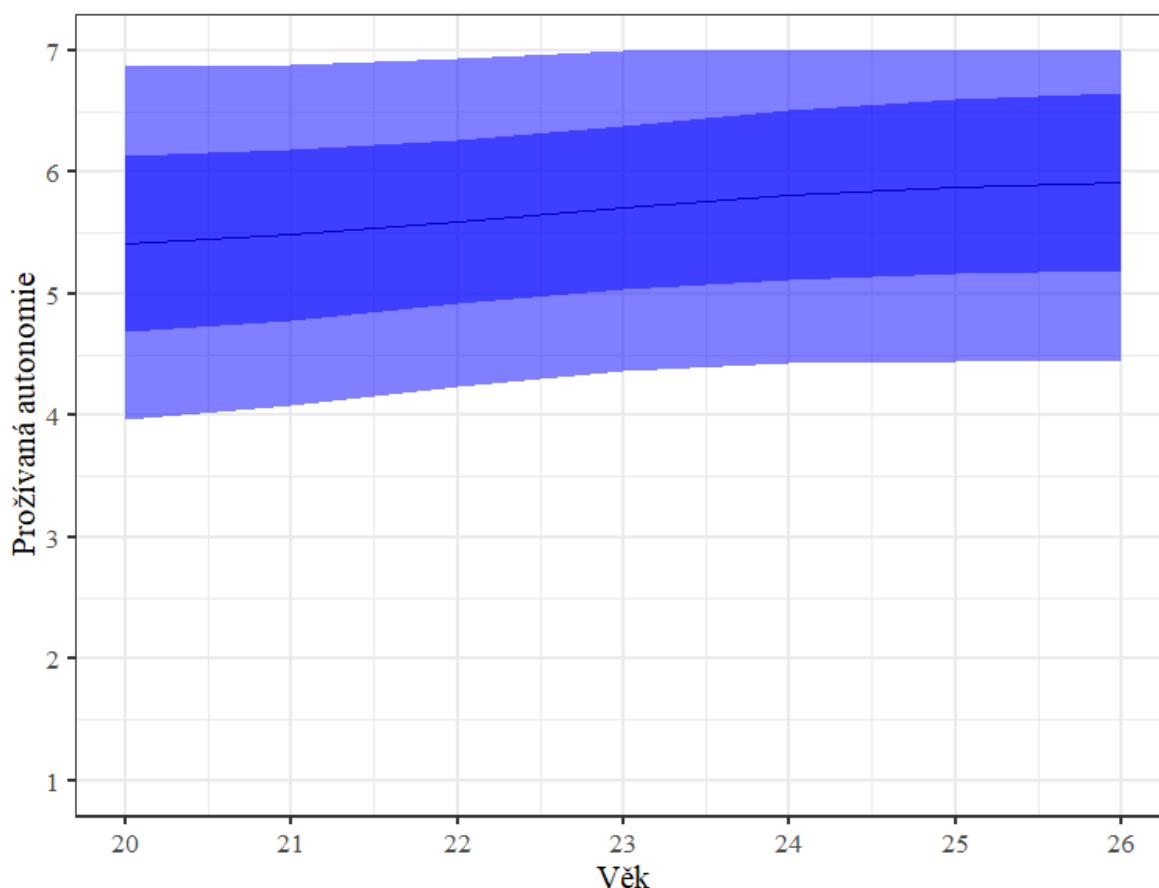
Tabulka IV.17. Parametry, odvozené statistiky a ukazatele shody CFM modelů.

	Lineární růst (M1)	Latent-base růst (M2)	Kvadratický růst (M3)	Logistický růst (M4)
<i>Růstové parametry</i>				
Průsečík (I)	5,41	5,42	5,38	5,33
Směrnice (Věk-20) (S)	0,10	0,08	0,12	0,63
Časové báze		0-0,7-2,1-3,6- 4,6- 5,7-6	0 – 6 0 – 36	0,13-0,24-0,41- 0,60-0,77-0,88- 0,94
Kvadr. směrnice (Věk-20) ² (Q)			-0,01*	
Lambda				2,48
Alfa				0,78
<i>Rozptyly</i>				
Disturbance A20-A26	0,22	0,22	0,20	0,21
Průsečík (I)	0,54	0,52	0,50	0,59
Směrnice (Věk-20) (S)	0,01	0,01	0,02*	0,53
Kvadr. směrnice (Věk-20) ² (Q)			<0,001*	
COV (I; S)	-0,03	-0,03	-0,01*	-0,27
COV (I; Q)			-0,01*	
COV (S; Q)			-0,001*	
Korelace disturbancí	-0,02 – 0,43	-0,09 – 0,28	-0,12 – 0,22	-0,10 – 0,27
<i>Odvozené statistiky</i>				
r (I; S)	-0,42	-0,38	-0,11	-0,49
r (I; Q)			-0,84	
r (S; Q)			-0,59	
<i>Ukazatele shody s daty</i>				
Chi2	2000	1992	1991	1993
df	1005	1000	1001	1003
Volných parametrů	220	225	224	222
RMSEA	0,028	0,028	0,028	0,028
CFI	0,925	0,925	0,926	0,926
SRMR	0,070	0,071	0,070	0,070
AIC	95076	95078	95075	95073
BIC	96201	96228	96220	96208
SSABIC	95502	95513	95508	95503

Poznámka. Není-li parametr označen *, popř. není-li uvedeno jinak, $p < 0,01$.

Všechny čtyři CFM modely mají nejen podobnou shodu s daty, ale i velmi podobné predikce. Odhadnuté časové báze v latent-base modelu se příliš neodchylují od lineárního trendu – nejprve je růst mírně pomalejší, od 23. roku o něco málo rychlejší a po 25. roce se zase zpomaluje. Takový mírně sigmoidní trend odráží i model logistického růstu (M4). Ten parametrizuje růst výrazně odlišně od lineárního modelu, takže přímým pohledem na parametry modelu se predikce modelů srovnávají jen obtížně. Odhadnuté časové zde udávají, jaká část celkového růstu se k danému roku realizovala. Mezi 20. a 21. rokem se tedy v průměrné trajektorii realizuje 11 % růstu (0,24 – 0,11), zatímco mezi 22. a 23. se realizuje 19 % růstu a mezi 25. a 26. již jen 6 % růstu. V logistickém růstu jsou počátek a konec růstu (0 a 1) mimo modelované rozpětí, v nekonečnu, díky čemuž se může pro přesnější napasování logistické křivky na data využít jejich různých úseků. Kvadratický růst se v CFM modelu prožívané autonomie příliš neosvědčil – kvadratická směrnice má sice očekávanou zápornou hodnotu odpovídající zpomalujícímu se růstu. Její průměrná hodnota -0,01 se však významně neliší od nuly a ani nemá významně nenulový rozptyl. Spolu s ním klesl pod hladinu významnosti i rozptyl směrnic a zřetelně narostly korelace mezi růstovými parametry. I přes dobrou shodu s daty nemá kvadratický model potenciál dále vysvětlit heterogenitu růstových křivek, protože jejich parametry nejsou zjevně dostatečně přesně odhadnuty.

Graf IV.15. Predikované růstové trajektorie logistickým CF modelem, pásmo +/- 2SD



Pro další modelování heterogenity růstových křivek pomocí mixture modelů jsem zvolil model logistického růstu. Statisticky jej lze považovat za rovnocenný lineárnímu modelu, který je jednodušší. Oproti lineárnímu modelu má dva parametry navíc a ukazatele penalizující za počet parametrů by preferovaly lineární model (SABIC, BIC). Ostatní ukazatele

jsou stejné nebo ve prospěch logistického růstu. Lineární růst je však teoreticky hůře obhajitelný, protože vývoj se zde odehrává blízko horní hranice rozpětí měřítka prožívané autonomie. Navíc, řada psychologických proměnných se vyvíjí spíše tak, že jsou kratší období rychlejšího růstu střídána delšími obdobími stability, což také více odpovídá logistickému růstu. Také z didaktických důvodů se zdá logistický růst zajímavější. Oproti latent-base modelu je logistický růst při větším počtu měření parametricky úspornější, dává však větší prostor pro odlišnosti tvaru růstových křivek v případných latentních třídách, jak tomu bylo u vývoje rizikového chování v předchozích kapitolách. Množství parametrů je zde již velmi vysoké, a tak má volba zůstala u modelu logistického růstu. Heterogenitu jím predikovaných individuálních růstových křivek zobrazuje graf na obrázku IV.15 zachycující pásmo dvou směrodatných odchylek nad a pod průměrnou trajektorií. Horní mez růstu je hodnota 7 na škále prožívané autonomie, proto je graf v pravém horním rohu „oříznutý“.

MODEL LATENTNÍCH RŮSTOVÝCH TŘÍD CFM

V modelu logistického růstu faktoru prožívané autonomie je modelována nezanedbatelná variabilita individuálních trajektorií vývoje autonomie. Příčiny individuálních rozdílů lze i zde hledat ve známých proměnných nebo v neznámých, latentních, subpopulacích, které se liší svým vývojem, přičemž ještě není jasné proč. I u růstových modelů druhého řádu lze hledat latentní třídy v pravém slova smyslu – tedy vnitřně zcela homogenní subpopulace, s nulovým rozptylem růstových parametrů – a latentní komponenty směsi – tedy subpopulace s nenulovou vnitřní variabilitou. Podobně jako v předchozích analýzách se nejprve pokusím zachytit heterogenitu růstových trajektorií pomocí vnitřně homogenních latentních tříd (SO-LCG, second-order latent class growth) a poté pomocí vnitřně heterogenních tříd, tedy mixture modelů (SO-GMM, second-order growth mixture models).

Specifikoval jsem dva druhy modelu latentních růstových křivek faktorů. V jednom je tvar růstové křivky shodný napříč latentními třídami a třídy se liší pouze v latentním průsečíku a směrnici, tedy v počátečním stavu před započítáním projektu a celkové míře růstu. V druhém se parametry určující tvar logistického růstu mohou lišit napříč latentními třídami (alfa a lambda), třídy se tedy ještě navíc mohou lišit v tom, v jakém čase je tempo růstu nejvyšší a jak strmý je růst.

Problémy s odhadem

Modely latentních tříd nad CFM modelem jsou výpočetně velmi náročné a obvykle se potýkají s konvergenčními problémy. Již u modelů se dvěma vnitřně homogenními latentními třídami se čas odhadu modelu začal pohybovat v hodinách na moderním dvanáctijádrovém procesoru, a to i při poměrně konzervativním počtu sad náhodných startovacích hodnot (např. STARTS = 200 40). Při navyšování počtu tříd čas roste natolik, že po hodinách běhu nezřídka výpočetní proces Mplus skončí, aniž by program vypsal jakýkoliv výstup či byla zaznamenána chyba. Protože podstatnou část odhadovaných parametrů tvoří parametry modelu měření, které by měly být stejné bez ohledu na to, kolik latentních tříd hledáme, nabízí se řešení v podobě převzetí hodnot těchto parametrů z CFM modelu. To může mít dvě podoby. Lze je napevno zafixovat na jejich odhady z CFM modelu, nebo je použít jako startovací hodnoty. Pevné zafixování vede k mírně horší shodě výsledného modelu latentních tříd s daty, protože tyto odhady jsou zatíženy chybou, která se v růstných modelech projeví různě. Čas potřebný na výpočet poklesne řádově. Zásadním problémem zde je ovšem velký pokles počtu volných

parametrů, který figuruje v mnoha ukazatelích shody modelu s daty. Ty jsou pak zkrácené a je nutné je ručně přepočítat. Použití odhadů parametrů modelu měření jako startovacích hodnot se zdá být vhodnější. Vede k lépe pasujícímu modelu díky tomu, že jejich hodnota se během iterací může ještě mírně změnit. Ukazatele shody modelu s daty nejsou zkrácené a časová úspora je jen mírně nižší než v případě fixování. Zároveň je výsledek takového postupu identický s výsledkem prostého odhadu parametrů bez použití startovacích hodnot (nenastanou-li zde konvergenční problémy).

Mplus umožňuje uživatelům pohodlné využití odhadnutých parametrů jako startovacích hodnot. Pomocí příkazu `SVALUES`; v sekci `OUTPUT`: se do výstupu zahrne syntax pro aktuální model s odhadnutými hodnotami parametrů jako startovacími hodnotami. Ten lze jednoduše zkopírovat do sekce `MODEL`: v nové analýze. Výhodou je i standardně vyšší počet desetinných míst, než je ve výpisu odhadů parametrů.

Postup s využitím odhadů parametrů z CFM modelu jako startovacích hodnot parametrů v SO-LCG modelu jsem použil i v níže popsanych analýzách. I přesto se při narůstajícím počtu tříd časy odhadu modelů šplhaly do řádu hodin, a proto je v tabulkách pro orientaci uvádím. Tato výpočetní náročnost má ještě jednu nepříjemnou konsekvenci, a to tu, že téměř znemožňuje BLR test pro testování toho, zda je model s více třídami lepší než model s méně třídami. Ten čas potřebný na výpočet posouvá do řádu dnů, a to ještě s nejistým výsledkem. Proto v dále uvedených analýzách spoléhám na BIC, který v předchozích analýzách a i podle dosavadních publikovaných zkušeností nejčastěji koresponduje s výsledkem BLR testu (Nylund et al., 2007).

SO-LCG model se stejným tvarem logistického růstu napříč třídami

Tabulka IV.18 prezentuje ukazatele shody s daty pro modely s postupně narůstajícím počtem latentních tříd. Pro každou třídu se odhadují dva růstové parametry – latentní průsečík a latentní směrnice. Ty jsou konstantami s nulovým rozptylem uvnitř tříd. Všechny ostatní parametry jsou napříč třídami identické. Pro každou třídu se také odhaduje její relativní četnost; počet volných parametrů modelu tak roste o 3 s každou latentní třídou navíc. Protože část variability se díky modelování růstu latentního skóru autonomie přesunula do reziduální části modelu, je zde méně heterogenity a méně prostoru pro latentní třídy. Proto, a i s ohledem na problémy odhadu a výpočetní čas, je maximální počet tříd 5.

Tabulka IV.18. Souhrnné ukazatele SO-LCG modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami. Na prvním řádku pro srovnání výchozí CFM model.

Počet tříd	par	LL	AIC	BIC	aBIC	AICC	Entropie	čas
CFM	222	-47315	95073	96208	95503			
2	222	-47373	95190	96325	95619	95289	0,713	4:59
3	225	-47298	95046	96196	95482	95148	0,678	26:11
4	228	-47276	95008	96173	95449	95113	0,731	1:03:56
5	231	-47252	94966	96147	95413	95074	0,748	1:51:24

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. AICC počítán balíčkem MplusAutomation (Hallquist & Wiley, 2018)

Tabulka IV.19. Parametry SO-LCG modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami.

		C	2	SE	3	SE	4	SE	5	SE	
Relativní četnosti		1	0,25		0,42		0,50		0,41		
tříd		2	0,75		0,52		0,07		0,47		
		3			0,06		0,01		0,01		
		4							0,005		
		5							0,10		
	Průměry	I	1	4,48	0,09	4,95	0,10	6,42	0,05	5,14	0,07
růstových	S	1	0,49	0,12	0,61	0,14	-0,42	0,07	0,44	0,09	
parametrů	I	2	5,77	0,06	5,95	0,07	4,15	0,11	6,03	0,06	
	S	2	0,49	0,10	0,51	0,10	-0,04	0,15	0,43	0,07	
	I	3			3,95	0,13	6,42	0,24	2,59	0,44	
	S	3			0,24	0,18	-3,68	0,56	3,85	0,57	
	I	4					5,47	0,07	3,59	0,61	
	S	4					-0,40	0,08	-1,47	0,77	
	I	5							4,26	0,11	
	S	5							0,17	0,13	
	Reziduální rozptyly	A20-A26	všechny	0,35	0,02	0,28	0,02	0,26	0,01	0,25	0,01
	Časové báze	A20	všechny	0,05	0,05	0,08	0,06	0,97	0,03	0,04	0,04
A21		všechny	0,15	0,08	0,18	0,09	0,88	0,06	0,15	0,07	
A22		všechny	0,34	0,09	0,37	0,09	0,62	0,06	0,42	0,07	
A23		všechny	0,60	0,08	0,61	0,07	0,27	0,08	0,75	0,07	
A24		všechny	0,82	0,09	0,81	0,08	0,08	0,06	0,93	0,05	
A25		všechny	0,93	0,06	0,92	0,06	0,02	0,02	0,98	0,02	
A26		všechny	0,98	0,03	0,97	0,03	0,00	0,01	1,00	0,01	
Alfa		všechny	1,09	0,36	0,98	0,31	-1,48	0,42	1,425	0,42	
Lambda	všechny	2,63	0,29	2,55	0,30	2,32	0,18	2,235	0,18		

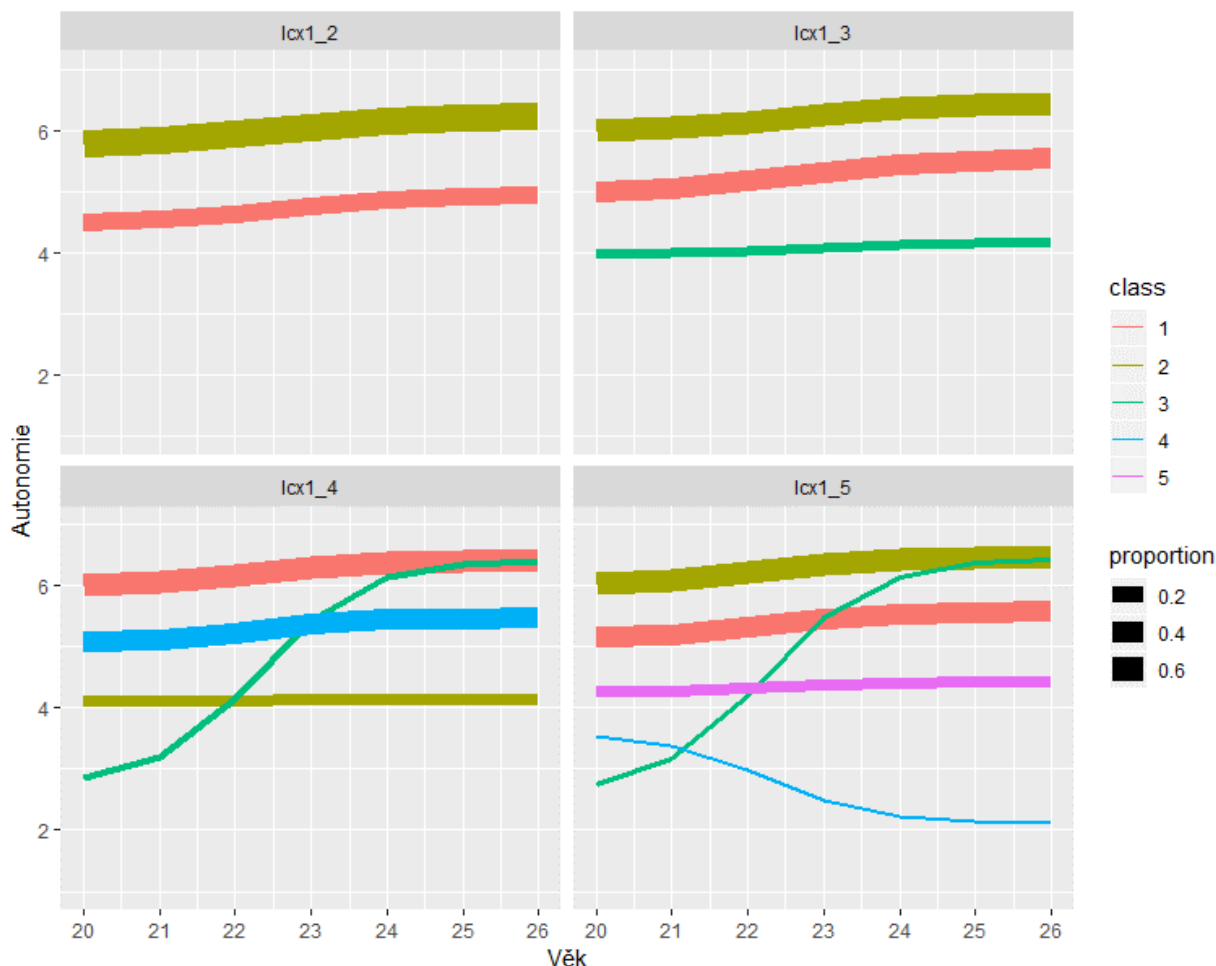
Poznámka. I - intercept, počáteční stav, S - směrnice, tempo růstu za rok, C - latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Časové báze pro jsou vypočítány z odhadnutých parametrů logistické růstové křivky - alfy a lambdy. Čas je z hlediska těchto parametrů kódovaný jako (Věk-20).

Je zajímavé, že oproti mixture modelům na manifestních proměnných prezentovaných výše, vykazuje model latentních tříd již se třemi třídami lepší shodu s daty než výchozí LGC model. Je to překvapivé, protože veškerá heterogenita je v modelu s latentními třídami redukována na fixní hodnoty růstu uvnitř tříd - všechny rozptyly je tak reprezentován pouze rozdíly mezi třídami.

Podle všech ukazatelů shody modelu s daty se jeví být nejlepší model s 5 třídami. Pohled na parametry jednotlivých modelů (tabulka IV.19) ukazuje, že již od modelu se třemi třídami se objevují velmi malé třídy, které mohou být ve vzorku reprezentovány jen několika jednotlivci. Jejich validita a replikabilita je tak velmi sporná. Třídou 4 v modelu s pěti třídami reprezentuje jen 6 účastníků, pro které je pravděpodobnost členství v této třídě nejvyšší. Zdá se tedy, že navyšování počtu tříd pouze umožňuje hypotetizování tříd, do kterých by patřili jednotlivci, kteří příliš nezapadají do majoritních tříd. Této představě také odpovídá to, že od

modelu se 3 třídami k modelu s 5 třídami postupně narůstá entropie. Další navyšování počtu tříd by tak možná vedlo k dalšímu nárůstu shody modelu s daty, ale ne k vynoření jiné struktury tříd.

Obrázek IV.16. Růstové trajektorie jednotlivých latentních tříd v SO_LCG modelech se 2 až 5 latentními třídami



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Rozptýl růstových křivek uvnitř tříd není znázorněn. Barvy jsou arbitrární - stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

Pro lepší představu o růstové křivce v jednotlivých latentních třídách jsou všechny modely zobrazeny v grafech na obrázku IV.16. Zde je také vidět, do jaké míry ovlivňují podobu růstových křivek parametry I a S . Všechny křivky v rámci jednotlivých panelů na obrázku IV.16 totiž mají stejné parametry α a λ . Ani napříč panely/modely se však tyto dva parametry příliš neliší. Oblast nejrychlejšího růstu je mezi 22. a 23. rokem ($20 + \lambda$) a tempo růstu reprezentované parametrem α nabývá hodnot mezi 1,0 a 1,5. Latentní směrnice S udává, jak moc růstu se v té které latentní třídě uskutečňuje. Například v první třídě v modelu se dvěma třídami má S hodnotu 0,49, což je rozdíl mezi autonomií na počátku a na konci sledovaného období.

Celkově jsou si růsty ve všech třídách, které reprezentují alespoň 5 % populace, velmi podobné. Mírně stoupající trend, přičemž střední třída s počáteční hodnotou kolem hodnoty 5 roste velmi mírně rychleji než třídy s počátkem níže či výše. I když modely se 4 a 5 třídami zahrnují i výrazně rychlejší růst a pokles, jde o velmi málo početné třídy, u nichž není jisté,

zda nejsou pouze artefaktem datového souboru. I přes tyto malé třídy se zde zdá být model s pěti latentními třídami nevhodnější, protože se solidní pravděpodobností klasifikuje většinu vzorku a se členy velmi malých tříd umožňuje zacházet podle teoretické orientace uživatele takové klasifikace (vyřadit, zahrnout do interpretací a dalších analýz).

SO-LCG model s různým logistickým růstem napříč třídami

Protože vlastně není důvod hypotetizovat, že by růst měl mít stejnou podobu napříč růstovými třídami, lze předchozí sadu modelů vnímat jako zjednodušený model vycházející vstříc výpočetním nárokům. Proto jsem se pokusil odhadnout také modely, v nichž by se parametry logistického růstu alfa a lambda mohly napříč skupinami lišit. Pro každou latentní třídu se tak odhaduje průměrný/konstantní latentní průsečík a latentní směrnice a alfa s lambdou – s relativní četností třídy je to dohromady 5 parametrů na každou další třídu. Ukazatele shody s daty pro modely s postupně narůstajícím počtem latentních tříd jsou uvedeny v tabulce IV.20. I zde jsem měl v úmyslu odhadnout model až s pěti latentními třídami. Ten již se však nepodařilo odhadnout – jeho odhad skončil při několika pokusech po více než 6 hodinách opadnutím zátěže procesoru bez jakékoli chybové hlášky. Přitom model se 4 třídami vykazuje lepší informační ukazatele než předchozí modely, a tak by případný model s 5 třídami stálo za to mít odhadnutý. Vzhledem k tomu, že v předchozí sadě modelů měl model s 5 třídami již dvě velmi malé třídy, lze soudit, že pro tak malé třídy mohly být parametry logistické růstové křivky empiricky neidentifikované.

Tabulka IV.20. Souhrnné ukazatele SO-LCG modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami. Na prvním řádku pro srovnání výchozí CFM model.

Počet tříd	par	LL	AIC	BIC	aBIC	AICC	Entropie	čas
CFM	222	-47315	95073	96208	95503			
2	224	-47373	95193	96338	95627	95294	0,712	15:45
3	229	-47297*	95051	96222	95494	95157	0,676	1:25:32
4	234	-47269*	95007	96203	95459	95118	0,709	4:08:51
5	239	---						

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. AICC počítán balíčkem MplusAutomation (Hallquist & Wiley, 2018). *Odhad vedl k nepřijatelnému řešení.

Srovnáme-li shodu těchto modelů s daty se SO-LCG modely s konstantními parametry logistického růstu napříč latentními třídami je zřejmé, že modely se stejným počtem tříd odpovídají datům v podobné míře. Vzhledem k tomu, že na to potřebují o dva parametry na třídu více, všechna informační kritéria penalizující za počet parametrů vychází hůře. Tato „přeparametrizovanost“ modelu se projevuje i v samotných odhadech růstových parametrů. Zatímco v modelu se dvěma třídami je ještě vše v pořádku a alfa a lambda se v obou třídách pohybují okolo hodnot, které měly, když musely být pro obě třídy stejné, již v modelu se třemi třídami došlo k tomu, že v nejmenší latentní třídě nejsou tyto dva parametry dobře empiricky identifikované. Alfa má hodnotu -36 (S.E. = 543) a lambda zůstala na startovací hodnotě 3⁹. Přitom se 7 % tato třída ještě nepatří mezi ty, kterou bych obecně považoval za příliš malou.

⁹ Zafixována iteračním algoritmem, aby se zabránilo singularitě informační matice.

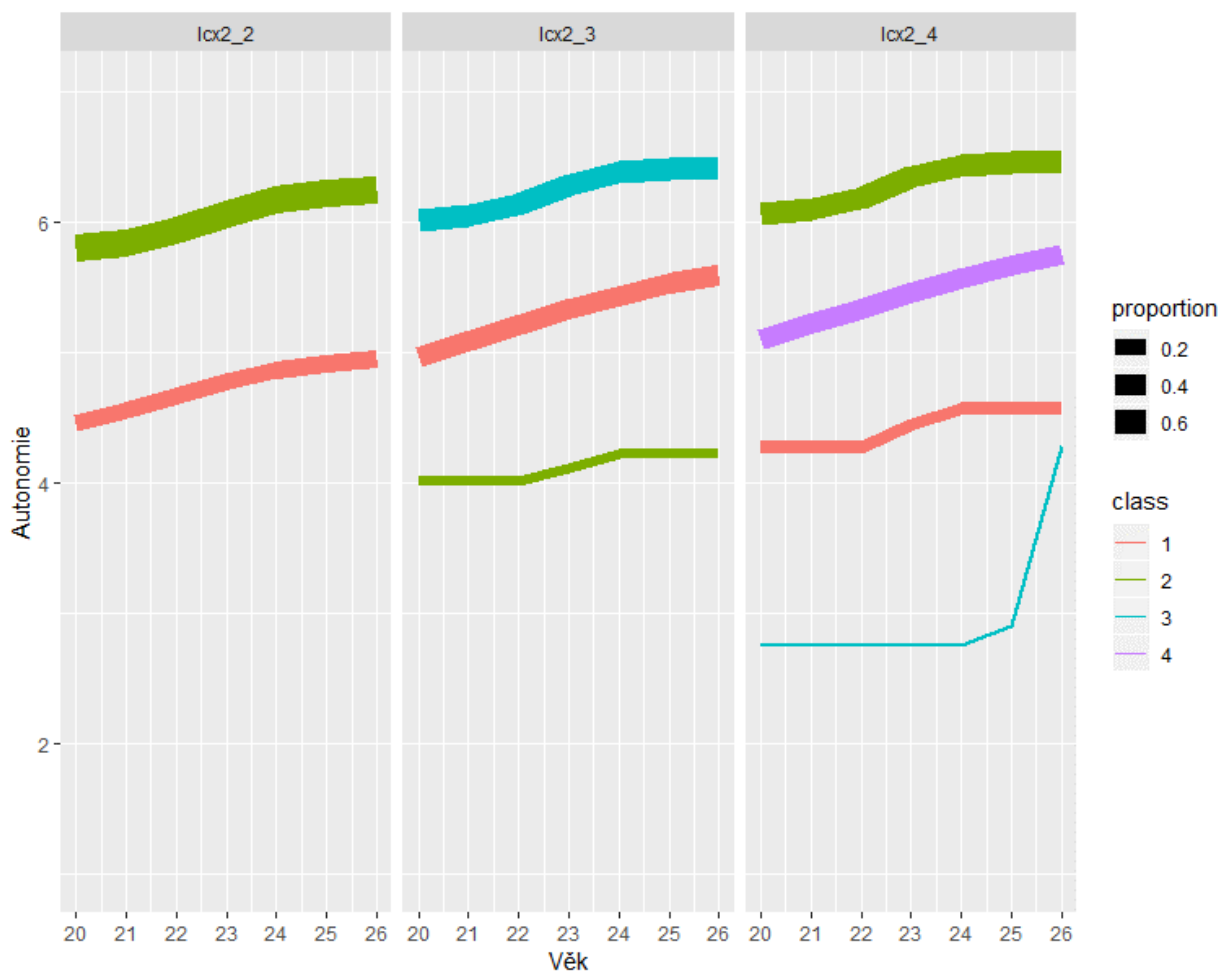
Podobně dopadly dvě malé třídy v modelu se čtyřmi latentními třídami – v obou je tvar růstu nutné považovat za v podstatě nezjištěný. Na druhou stranu se zde neobjevuje nic, co by se výrazně lišilo od tříd identifikovaných předchozí sadou modelů, což vypovídá o tom, že rozdílně specifikované modely nalézají v datech tytéž pravidelnosti.

Tabulka IV.21. Parametry SO-LCG modelů prožívané autonomie založených na modelu logistického růstu s proměnlivým tvarem napříč třídami.

		C	2	SE	3	SE	4	SE
Relativní četnosti		1	0,25		0,42		0,11	
tříd		2	0,75		0,07		0,46	
		3			0,51		0,01	
		4					0,42	
Průměry	I	1	4,98	0,16	4,52	1,38	4,27	0,09
růstových	S	1	-0,65	0,66	1,24	1,77	0,29	0,10
parametrů	I	2	6,26	0,07	4,21	0,13	6,46	0,05
	S	2	-0,48	0,10	-0,21	0,12	-0,41	0,08
	I	3			6,01	0,06	4,34	7,49
	S	3			0,41	0,08	-1,58	7,52
	I	4					4,48	2,51
	S	4					1,63	3,57
Reziduální rozptyly	A20-A26	všechny	0,35	0,02	0,28	0,02	0,26	0,01
Parametry	Alfa	1	-0,72	0,86	0,40	0,55	63,51	0,00
logistického	Lambda	1	1,90	2,25	1,46	4,32	3,00	0,02
růstu	Alfa	2	-1,12	0,43	-35,86	543,17	-1,62	0,65
	Lambda	2	2,71	0,35	3,00	0,00	2,46	0,29
	Alfa	3			1,54	0,62	-5,69	145,74
	Lambda	3			2,51	0,29	5,40	11,06
	Alfa	4					0,29	0,64
	Lambda	4					1,68	6,56

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Časové báze lze vypočítat z parametrů alfa a lambda jako $1/(1 + \text{EXP}(-((\text{Věk}-20)-\text{lambda})*\text{alfa}))$. Čas je z hlediska těchto parametrů kódovaný jako (Věk-20).

Obrázek IV.17. Růstové trajektorie jednotlivých latentních tříd v SO_LCG modelech se 2 až 4 latentními třídami



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Rozptyl růstových křivek uvnitř tříd není znázorněn. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

V případě SO-LCG modelů autonomie tedy data neposkytují dostatek informace pro určení tvaru křivky v jednotlivých latentních třídách. Modely s konstantním tvarem růstové křivky napříč třídami jsou proto vhodnějším popisem dat.

MODEL RŮSTOVÝCH SMĚSÍ CFM – SO-GMM

Na rozdíl od modelů latentních růstových tříd umožňují mixture modely, aby růstové parametry měly nenulovou variabilitu růstových parametrů uvnitř latentních tříd. Z hlediska specifikace modelu s latentními třídami to znamená, že je potřeba se rozhodnout, zda umožnit či neumožnit třídám, aby se lišily v rozptylech růstových parametrů. Umožnit různé rozptyly růstových parametrů napříč třídami znamená další uvolnění modelu, který jsem již výše označil za přeparametrizovaný, a proto jsem se rozhodl SO-GMM modely specifikovat s konstantními rozptyly růstových parametrů napříč třídami. Specifikoval jsem opět dva druhy SO-GMM. V obou je tvar růstové křivky shodný napříč latentními třídami. V tom prvním se třídy liší pouze v latentním průsečíku a směrnici, tedy v počátečním stavu před započítáním projektu a celkové míře růstu. V druhém se třídy mohou lišit i v disturbancích a

jejich kovarianci. To je analogické GMM modelům s manifestními proměnnými, v nichž se třídy mohly lišit svou reziduální strukturou. Oba modely vycházejí z CFM modelu s logistickým růstem a pro měřící část modelu používají startovací hodnoty parametrů z výchozího CFM modelu.

SO-GMM model s třídami lišícími se pouze průměry růstových parametrů

Protože i rozptyly růstových parametrů jsou v tomto modelu konstantní napříč třídami, každá třída navíc navyšuje počet volných parametrů pouze o tři – průměry latentního průsečíku a směrnice a relativní četnost třídy. Z hlediska BIC se jeví být optimálním modelem ten se čtyřmi latentními třídami. Neliší se však příliš od modelu se třemi třídami, přičemž i entropie se u obou modelů podobá. Oba modely obsahují alespoň jednu velmi malou třídu. V tom navazují na model se dvěma třídami, který dělí populaci v poměru 9:1. Je tedy vidět, že výchozí CFM model popisuje dobře většinu populace a postupně se vynořující latentní třídy jsou spíše minoritními, okrajovými subpopulacemi. Tyto třídy reprezentují spíše heterogenitu tempa růstu, o čemž svědčí to, že právě rozptyl latentní směrnice je ten, který v modelu se třemi třídami ztlačil oproti modelu se dvěma třídami; rozptyl latentních průsečíků zůstal stejný i přes nárůst počtu skupin. Ani v modelu se dvěma třídami, ani v následujících modelech se rozptyl latentních směrnic neliší signifikantně od nuly. Postupně s nárůstem počtu latentních tříd oba rozptyly klesají a při pěti třídách jsou téměř nulové a model se tak stává téměř SO-LCG modelem. Iterační algoritmus zde již postupoval velmi pomalu, o čemž svědčí skokový nárůst doby potřebné pro odhad parametrů modelu. Přesto jsem se při odhadu parametrů této sady modelů setkal s méně konvergenčními problémy než u SO-LGC modelu s různými růstovými křivkami napříč třídami.

Tabulka IV.22. Souhrnné ukazatele SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami. Na prvním řádku pro srovnání výchozí CFM model.

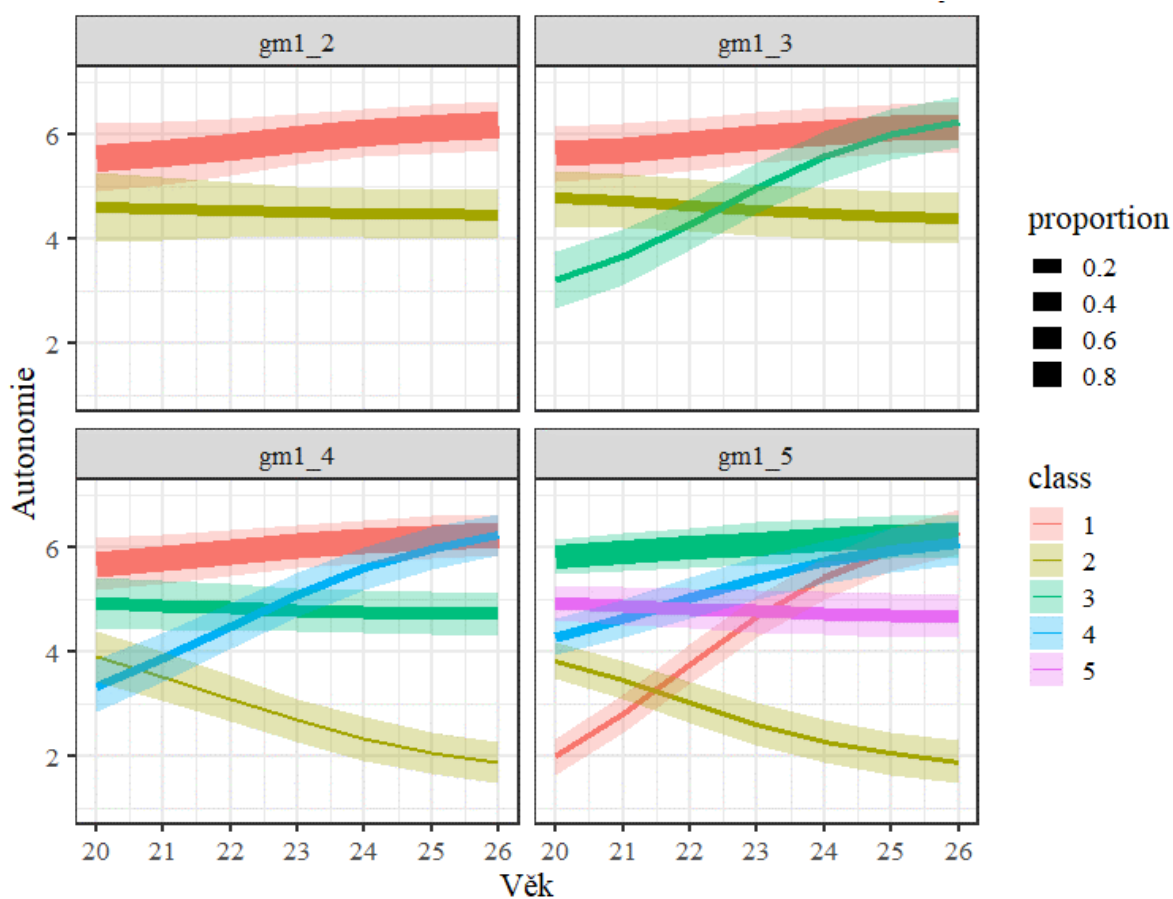
Počet tříd	par	LL	AIC	BIC	aBIC	AICC	Entropie	čas
CFM	222	-47315	95073	96208	95503			
2	225	-47281	95012	96162	95447	95113	0,710	6:53
3	228	-47260	94975	96141	95416	95080	0,751	38:07
4	231	-47248	94959	96139	95406	95066	0,740	25:24
5	234	-47241	94951	96147	95404	95062	0,695	1:42:08

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. AICC počítán balíčkem MplusAutomation (Hallquist & Wiley, 2018)

Rozhodování o nejvhodnějším modelu se tak zde odehrává mezi modelem se třemi třídami a modelem se čtyřmi třídami. Nápomocné v tom jsou i střední růstové křivky latentních tříd v jednotlivých modelech vykreslené na obrázku IV.18. Napříč modely zůstává majoritní třída (80–90 %) s vysokou počáteční mírou prožívané autonomie a mírným růstem a menší třída (10–15 %) se střední mírou prožívané autonomie, která neroste, či dokonce klesá. Právě z této menší třídy se v modelu se třemi třídami vyděluje malá třída (3–4 %) s nízkou počáteční mírou prožívané autonomie, ale rychlým růstem, jímž se na konci sledovaného období třída dostane na úroveň majoritní třídy. Toto vydělení téměř neovlivní parametry původních dvou tříd, pouze sníží rozptyl latentních směrnic. Oproti tomu třída, která se nově objevuje v modelu se 4 třídami, již způsobuje větší změny – majoritní třída se zmenšuje, menší

třída se zvětšuje, období nejprudších změn (λ) se přesouvá o tři čtvrtě roku dříve. To vše může souviset s tím, že v mladším věku máme k dispozici více dat a model se nově uvolněnými parametry přizpůsobuje více datům a roste pravděpodobnost artefaktuality zjištěných latentních tříd. Proto bych zde považoval za optimální model se třemi třídami – majoritní s vysokou rostoucí autonomií, minoritní se středně vysokou a stagnující autonomií a marginální „dohánějící“ třídou.

Obrázek IV.18. Růstové trajektorie jednotlivých latentních tříd v SO_GMM modelech se 2 až 5 latentními třídami



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Průsvitný pruh kolem křivky znázorňuje variabilitu růstových křivek uvnitř dané třídy – jednu směrodatnou odchylku nad a jednu pod průměrnou trajektorií. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

Tabulka IV.23. Parametry SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami.

		C	2	SE	3	SE	4	SE	5	SE
Relativní četnosti		1	0,90		0,88		0,80		0,01	
tříd		2	0,10		0,10		0,01		0,01	
		3			0,03		0,15		0,74	
		4					0,04		0,08	
		5							0,15	
Průměry	I	1	6,22	0,10	6,19	0,08	5,44	0,33	0,31	2,61
růstových	S	1	-0,79	0,20	-0,67	0,16	0,87	0,41	6,47	3,12
parametrů	I	2	4,44	0,15	4,37	0,17	4,74	1,66	4,58	1,44
	S	2	0,18	0,25	0,47	0,26	-3,15	2,27	-2,91	1,97
	I	3			6,50	0,40	5,02	0,21	5,65	0,25
	S	3			-3,82	1,17	-0,34	0,29	0,62	0,29
	I	4					2,14	1,77	3,57	1,16
	S	4					4,51	2,25	2,72	1,48
	I	5							5,03	0,21
	S	5							-0,39	0,27
Rozptyly	I	všechny	0,23	0,06	0,23	0,05	0,31	0,16	0,08	0,11
růstových	S	všechny	0,49	0,29	0,14	0,16	0,15	0,27	-0,07	0,21
parametrů	Cov(I,S)	všechny	-0,10	0,11	-0,03	0,08	-0,15	0,20	0,09	0,15
Reziduální rozptyly	A20-A26	všechny	0,21	0,02	0,21	0,02	0,21	0,02	0,21	0,02
Časové báze	A20	všechny	0,85	0,09	0,86	0,10	0,27	0,22	0,26	0,21
	A21	všechny	0,74	0,10	0,75	0,11	0,39	0,21	0,39	0,20
	A22	všechny	0,59	0,08	0,58	0,09	0,52	0,16	0,53	0,15
	A23	všechny	0,41	0,06	0,40	0,06	0,65	0,10	0,67	0,09
	A24	všechny	0,26	0,07	0,24	0,06	0,76	0,06	0,79	0,05
	A25	všechny	0,15	0,07	0,13	0,06	0,85	0,06	0,87	0,05
	A26	všechny	0,08	0,06	0,07	0,05	0,91	0,06	0,92	0,05
	Alfa	všechny	-0,71	0,24	-0,74	0,24	0,55	0,23	0,59	0,27
	Lambda	všechny	2,49	0,41	2,45	0,42	1,85	1,23	1,77	1,11

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Časové báze pro jsou vypočítány z odhadnutých parametrů logistické růstové křivky – alfy a lambdy. Čas je z hlediska těchto parametrů kódovaný jako (Věk-20).

SO-GMM model s třídami lišícími se průměry růstových parametrů a volnější reziduální strukturou

Předchozí modely byly poměrně silně svázané požadavkem na rovnost reziduí napříč třídami. To v zásadě odpovídá předpokladu, že krátkodobá dynamika prožívané autonomie je ve všech třídách stejná. Prožívaná autonomie je však jedním z aspektů prožívání, obsahuje emoční, pocitové prvky, a proto je na místě předpokládat, že se lidé ve stabilitě prožívané autonomie mohou lišit alespoň do určité míry podobně jako ve stabilitě emocí. Zároveň není důvod předpokládat, že by zastoupení lidí s kolísavějšími emocemi bylo napříč latentními třídami stejné. Proto jsem se pokusil odhadnout variantu předchozích modelů s tou změnou, že rezidua (tedy přesněji disturbance) latentní autonomie se mohou napříč třídami lišit. To dává prostor pro to, aby mohla být identifikována třída, které se nebude lišit tolik svou průměrnou růstovou křivkou, ale spíše krátkodobou variabilitou kolem individuálních růstových křivek, kterou rezidua reprezentují. Dává také prostor pro to, aby byli jednotlivci s více kolísajícím vývojem různě zastoupeni v různých latentních třídách. Je potřeba však připomenout, že rezidua nerepresentují pouze krátkodobé kolísání. Reprezentují také individuální vývoje, které svou formou nepasují na zvolený model. Třída s vysokými rezidui tak může být i třídou „misfitů“ - modelu neodpovídajících vývoju. I proto znamená uvolnění reziduí napříč třídami značný nárůst problémů s konvergencí modelu, patrně v návaznosti na nedostatečnou identifikaci jednotlivých parametrů.

Disturbance jsou ve výchozím CFM modelu fixovány, aby měly stejnou hodnotu napříč věky. To znamená, že v každé třídě se bude odhadovat jeden parametr navíc. Mají-li se lišit disturbance napříč třídami, musí se mít možnost lišit také kovariance mezi nimi, pokud nejsou fixovány na nulu. Ve výchozím CFM modelu jsou povoleny kovariance sousedících disturbance (autokorelační struktura), které se mohou navzájem lišit. Autokovariancí mezi sedmi věky je 6, což znamená 6 dalších parametrů navíc pro každou třídu. Celkově, spolu s průměry latentních růstových parametrů a velikostí třídy je pro každou další třídu v tomto modelu odhadováno 10 parametrů navíc.

Možná i kvůli této velké volnosti a kvůli velkému podílu chybějících dat měly tyto modely velké obtíže zkonvergovat do přípustného řešení. Vlastně ani jeden z modelů neposkytl zcela přípustné řešení. V modelech se dvěma a třemi třídami nebyla vždy v jedné ze tříd pozitivně definitní matice Ψ - tedy kovarianční matice latentních proměnných reprezentující autonomii. Šlo negativní reziduální rozptyl latentní proměnné v některém věku. Odhad těchto modelů probíhal překvapivě rychle (viz tabulka IV.24, ale bylo to bohužel díky tomu, že konvergenční problémy nastávaly velmi záhy. V modelu se čtyřmi a pěti latentními třídami dokonce Mplus v průběhu odhadu zafixoval řadu parametrů, aby se vyhnul singularitě informační matice, což zahrnovalo i zafixování relativní četnosti jedné ze tříd na 0. Efektivně se tak odhadoval model s méně třídami. Taková řešení jsou stěží přijatelná.

Vzhledem k výše uvedeným problémům s konvergencí a přípustností finálních odhadů parametrů jsou modely s vyšším množstvím tříd neakceptovatelné. Zaměřil jsem se tedy pouze na modely se dvěma třemi třídami. I tyto modely však nenabízí příliš použitelná řešení. V první řadě mají oba nízkou hodnotu entropie. Oddělení latentních tříd je tak nezřetelné a kategorizace jednotlivců by byla zatížena velkou nejistotou. Je to daň za to, že model může třídit lidi jak podle tvaru jejich růstové křivky, tak podle variability kolem této růstové křivky. Zde nastává situace, kdy ani jeden z těchto faktorů není zřetelně „třídotvorný“, což je zdrojem klasifikační nejistoty.

Tabulka IV.24. Souhrnné ukazatele SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami a lišící se reziduální strukturou. Na prvním řádku pro srovnání výchozí CFM model.

Počet tříd	par	LL	AIC	BIC	aBIC	AICC	Entropie	čas
CFM	222	-47315	95073	96208	95503			
2	232	-47040	94545	95731	94994	94654	0,554	20:35
3	242	-46998	94479	95716	94947	94599	0,521	16:56
3a	238	-47025	94526	95743	94987	94642	0,657	26:46
4	252	-46998	94499	95787	94987	94630	0,620	26:47
5	262	-46987	94498	95837	95005	94641	0,614	22:27

Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. AICC počítán balíčkem MplusAutomation. Všechny modely měly alespoň nějaký problém s nepřipustností konečných parametrů.

Pohled na parametry modelu do značné míry odhaluje, čím se algoritmus nechal vést. Ve všech modelech hledal alespoň jednu třídu s minimálními rezidui. Tím se dostával na hranu a v některých případech (např. třída 3 modelu se třemi třídami) byl odhadnutý rozptyl reziduí v této třídě negativní. Tuto třídu reprezentovali především respondenti s vysokou autonomií, která má tendenci vyvíjet se stabilněji, než střední a nižší úrovně prožívané autonomie. V modelu se dvěma třídami je to druhá třída a v modelu se třemi třídami již zmiňovaná třetí třída. Naopak algoritmus do značné míry ignoroval růstové parametry, což je zřejmé z toho, že s rostoucím počtem tříd rozptyl latentních průsečíků a směrnic příliš neklesá. Pokud bychom si připomněli základní cíl růstové mixture analýzy, tedy vysvětlení heterogenity růstu, pak nám aktuální model nic nepřináší. Zaměřuje se především na reziduální strukturu. Model se třemi třídami tak rozlišuje latentní třídy především podle reziduálního rozptylu – v první třídě vysokých 0,78, v druhé 0,18 a ve třetí nepřipustných -0,05. Standardizováno to znamená, že v první třídě vysvětlují růstové křivky méně než 20 % rozptylu latentních skóre autonomie, v druhé asi 50 % a ve třetí nepřipustných více než 100 %.

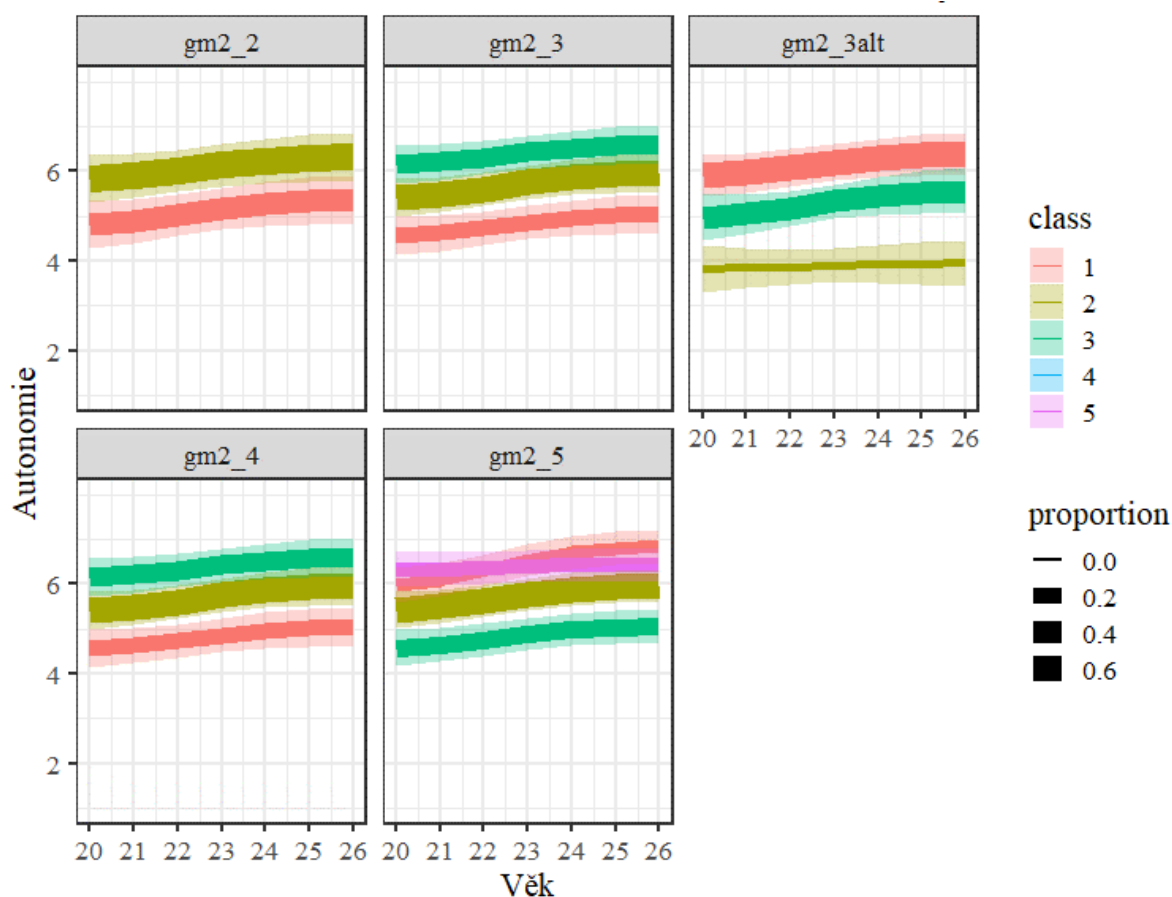
Celkově je třeba tuto variantu mixture růstového modelu považovat za neuspokojivou. Přesto jsem se ještě pokusil elaborovat model se třemi třídami a specifikoval jsem jeho alternativní verzi, v níž měly dvě třídy stejnou reziduální strukturu a jedna třída odlišnou – tedy hybridní specifikaci latentních tříd. Cílem bylo umožnit modelu utvořit jednu třídu s velkým reziduálním rozptylem a umožnit tak vynoření „čistších“ dvou tříd. Těmto dvěma třídám bylo také umožněno, aby se do té první lišily v rozptylu růstových parametrů. Model je v tabulkách označen 3a. Oproti původnímu modelu se třemi třídami má skokově vyšší entropii, ale jeho shoda s daty je mírně horší. Třídy, které tento model navrhuje, zahrnují opět velkou třídu s minimálními rezidui, vysokou mírou autonomie a nízkou variabilitou růstových parametrů a pak dvě třídy s většími rezidui a větším rozptylem růstových parametrů. Tento model je patrně reálnějším popisem dat.

Tabulka IV.25. Parametry SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s konstantním tvarem napříč třídami a lišící se reziduální strukturou.

		C	2	SE	3	SE	3a	SE	4	SE	5	SE	
Relativní četnosti tříd		1	0,38		0,19		0,54		0,19		0,08		
		2	0,62		0,54		0,03		0,54		0,58		
		3			0,27		0,43		0,27		0,20		
		4							0,00		0,00		
		5									0,13		
Průměry růstových parametrů	I	1	4,75	0,10	4,50	0,13	5,84	0,08	4,50	0,13	5,87	0,17	
	S	1	0,65	0,14	0,58	0,19	0,56	0,10	0,58	0,19	1,02	0,20	
	I	2	5,79	0,07	5,33	0,11	3,81	0,37	5,33	0,11	5,36	0,12	
	S	2	0,56	0,09	0,66	0,13	0,15	0,42	0,66	0,13	0,63	0,14	
	I	3			6,10	0,08	4,88	0,11	6,10	0,08	4,53	0,12	
	S	3			0,52	0,11	0,71	0,16	0,52	0,11	0,55	0,17	
	I	4							-3,91	0,00	1,79	0,00	
	S	4							-1,21	0,00	2,00	0,00	
	I	5									6,32	0,12	
	S	5									0,12	0,15	
	Rozptyly růstových parametrů	I	všechny	0,30	0,06	0,21	0,05	0,34	0,15	0,21	0,05	0,17	0,05
		S	všechny	0,33	0,13	0,30	0,13	0,69	0,37	0,30	0,13	0,12	0,09
Cov(I,S)		všechny	-0,17	0,08	-0,16	0,08	-0,37	0,28	-0,16	0,08	0,08	0,07	
I		1					0,25	0,06					
S		1					0,30	0,14					
Cov(I,S)		1					-0,16	0,08					
Reziduální rozptyly	A20-A26	1	0,57	0,04	0,78	0,08	0,01	0,01	0,78	0,08	0,04	0,02	
	A20-A26	2	0,02	0,01	0,18	0,03	0,46	0,04	0,18	0,03	0,18	0,03	
	A20-A26	3			-0,05	0,01	0,46	0,04	-0,05	0,01	0,79	0,07	
	A20-A26	4							0,05	0,00	0,05	0,00	
	A20-A26	5									0,07	0,02	
Časové báze	A20	všechny	0,10	0,06	0,11	0,06	0,11	0,06	0,11	0,06	0,10	0,06	
	A21	všechny	0,21	0,07	0,22	0,08	0,23	0,08	0,22	0,08	0,22	0,08	
	A22	všechny	0,41	0,06	0,41	0,06	0,42	0,07	0,41	0,06	0,43	0,07	
	A23	všechny	0,63	0,05	0,63	0,05	0,63	0,05	0,63	0,05	0,65	0,05	
	A24	všechny	0,81	0,05	0,81	0,06	0,80	0,05	0,81	0,06	0,83	0,05	
	A25	všechny	0,92	0,04	0,91	0,05	0,91	0,04	0,91	0,05	0,92	0,04	
	A26	všechny	0,97	0,03	0,96	0,03	0,96	0,03	0,96	0,03	0,97	0,02	
	Alfa	všechny	0,92	0,22	0,89	0,22	0,87	0,22	0,89	0,22	0,94	0,22	
	Lambda	všechny	2,41	0,24	2,40	0,25	2,37	0,26	2,40	0,25	2,33	0,25	

Poznámka. I - intercept, počáteční stav, S - směrnice, tempo růstu za rok, C - latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0. Časové báze pro jsou vypočítány z odhadnutých parametrů logistické růstové křivky - alfy a lambdy. Čas je z hlediska těchto parametrů kódovaný jako (Věk-20).

Obrázek IV.19. Růstové trajektorie jednotlivých latentních tříd v SO_GMM modelech se 2 až 5 latentními třídami



Poznámka. Tloušťka trajektorie vyjadřuje relativní četnost dané latentní třídy. Průsvitný pruh kolem křivky znázorňuje variabilitu růstových křivek uvnitř dané třídy – jednu směrodatnou odchylku nad a jednu pod průměrnou trajektorií. Barvy jsou arbitrární – stejná barva napříč modely nereprezentuje stejnou, analogickou třídu.

SO-GMM model s třídami lišícími se průměry růstových parametrů a parametry logistické křivky

Pokud se mají latentní třídy lišit i tvarem logistického růstu reprezentovaným parametry alfa a lambda, pro každou další třídu je potřeba odhadovat pět dalších parametrů – kromě alfy a lambda dva průměry latentních růstových parametrů a jedna relativní četnost třídy.

Tabulka IV.26. Souhrnné ukazatele SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s různým tvarem napříč třídami. Na prvním řádku pro srovnání výchozí CFM model.

Počet tříd	par	LL	AIC	BIC	aBIC	AICC	Entropie	čas
CFM	222	-47315	95073	96208	95503			
2	227	-47257	94967	96127	95406	95071	0,631	2:12:08
3	232	-47211	94886	96072	95335	94995	0,587	39:58
4	237	-47278	95029	96241	95488	95143	0,647	24:53
5	242	-47229	94942	96179	95410	95061	0,688	2:21:33

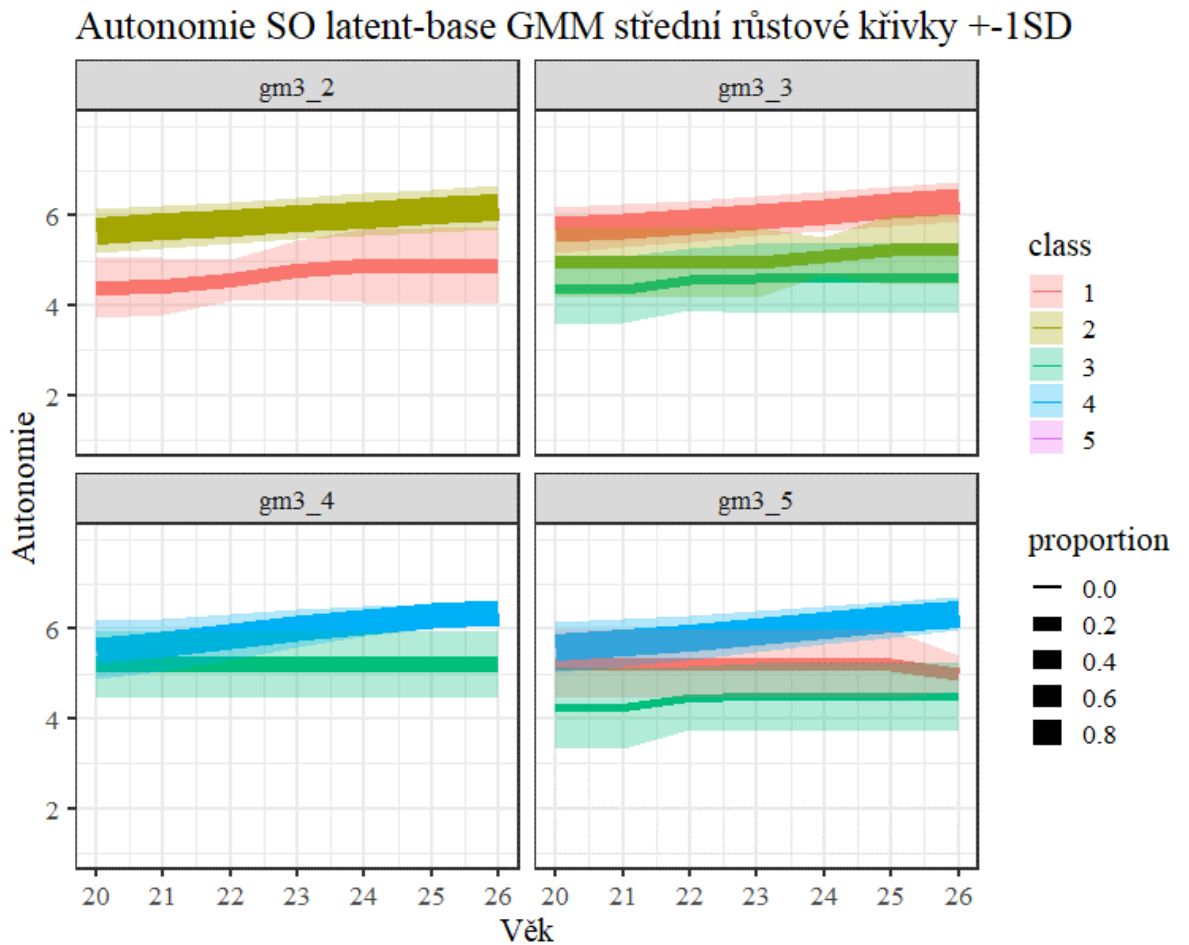
Poznámka. Tučně jsou vyznačeny nejlepší hodnoty daného parametru, popř. hodnota modelu, který by dle daného parametru byl optimální. AICC počítán balíčkem MplusAutomation.

Tabulka IV.27. Parametry SO-GMM modelů prožívané autonomie založených na modelu logistického růstu s různým tvarem napříč třídami.

		C	2	SE	3	SE	4	SE	5	SE
Relativní četnosti		1	0,16		0,81		0,00		0,14	
tříd		2	0,84		0,14		0,00		0,00	
		3			0,05		0,25		0,02	
		4					0,75		0,83	
		5							0,00	
Průměry	I	1	4,90	0,13	7,18	0,53	-3,56	0,00	4,85	0,22
růstových	S	1	-0,52	0,16	-2,13	0,74	1,93	0,00	0,36	0,24
parametrů	I	2	4,68	0,77	4,96	0,12	-4,97	0,00	-3,85	0,00
	S	2	2,12	1,35	0,28	0,16	-0,92	0,00	4,81	0,00
	I	3			4,60	0,16	-0,03	0,12	4,24	0,23
	S	3			-0,25	0,21	5,23	0,00	0,24	0,28
	I	4					6,44	0,12	4,66	0,44
	S	4					-1,10	0,22	3,66	1,15
	I	5							-3,84	0,00
	S	5							-1,47	0,00
Rozptyly	I	všechny	0,72	0,16	0,00	0,60	0,10	0,00	0,04	0,07
růstových	S	všechny	1,48	0,52	0,01	1,70	0,33	0,00	0,27	0,16
parametrů	Cov(I,S)	všechny	-0,86	0,26	0,00	-0,84	0,15	0,00	0,12	0,10
Reziduální rozptyly	A20-A26	všechny	0,72	0,16	0,00	0,60	0,10	0,00	0,04	0,07
Časové báze	Alfa	1	-1,91	0,77	-0,21	0,07	0,10	0,00	-16,11	443,33
	Lambda	1	2,27	0,32	4,15	1,75	2,68	0,00	5,98	0,67
	Alfa	2	0,18	0,12	60,97	0,00	0,61	0,00	5,39	0,00
	Lambda	2	0,94	2,72	4,00	0,01	3,73	0,00	6,53	0,00
	Alfa	3			-6,00	6,43	5,48	0,00	14,41	369,91
	Lambda	3			1,60	0,46	-4,73	0,00	1,78	5,56
	Alfa	4					-0,67	0,17	0,15	0,05
	Lambda	4					2,35	0,39	7,03	2,69
	Alfa	5							5,34	0,00
	Lambda	5							-2,93	0,00

Poznámka. I – intercept, počáteční stav, S – směrnice, tempo růstu za rok, C – latentní třída. Tučně jsou odhady parametrů a normálním řezem jsou jejich směrodatné chyby. Kurzívou jsou parametry, které se na 5% hladině neliší signifikantně od 0, červeně problematické parametry. Časové báze lze vypočítat z parametrů alfa a lambda jako $1/(1 + \text{EXP}(-((\text{Věk}-20)-\text{lambda}) \cdot \text{alfa}))$. Čas je z hlediska těchto parametrů kódovaný jako (Věk-20).

Obrázek IV.20. Růstové trajektorie jednotlivých latentních tříd v SO_GMM modelech se 2 až 5 latentními třídami



Shrnutí modelů autonomie

Souhrnná úvaha nad všemi prezentovanými modely heterogenity růstových křivek prožívané autonomie může vést k vícero závěrům. Tím prvním je ten, že model latentních růstových křivek (LGC) je poměrně dobrým popisem variability růstu, ať již v lineární nebo kvadratické podobě. Dvojnásob to platí pro růstový model druhé úrovně – CFM. Normální rozložení růstových koeficientů v těchto modelech dobře koresponduje s pozorovaným rozložením individuálních růstových trajektorií. Zde je však dobré si připomenout samotný začátek analýzy – pozorované růstové trajektorie – ty jsou v případě autonomie prototypickým nepřehledným spaghetti plotem. Lineární i kvadratický model reprezentují tyto individuální vývoje jen do té míry, do jaké je v nich přítomen lineární nebo kvadratický trend, a ta není až tak malá – normálně rozložené růstové křivky vysvětlují okolo dvou třetin rozptylu autonomie v každém věku. I v CFM modelu jsou vysvětlují latentní růstové křivky v různých specifikacích modelu kolem 70 % rozptylu latentních autonomií.

Modely latentních tříd růstových křivek, které snaží rozložení růstových křivek popsat několika diskretním třídami, nevychází oproti výchozím modelům lépe, a to ani v případě většího množství latentních tříd. Jejich přínosem v této analýze však bylo upozornění na důležitost reziduálních rozptylů. Vzhledem k tomu, že veškerá variabilita růstových křivek

uvnitř tříd je v těchto modelech právě v reziduálním rozptylu, facilitují tím uvažování o reziduích. V případě autonomie se napříč modely začala opakovaně objevovat třída s výrazně vyšším reziduálním rozptylem než v ostatních třídách. Poněkud jinak to bylo v případě SO-LCG modelů, kde již pomocí tří tříd bylo možná popsat heterogenitu latentních růstových křivek stejně či lépe, než to činil výchozí CFM model. Na druhou stranu latentní třídy měly paralelní průběh svých růstových křivek, takže vlastně neposkytovaly věcně jiný obraz než CFM model. V této fázi má tento obraz víceméně paralelních mírně rostoucích vývojů lišících se hlavně počáteční hodnotou autonomie, a jen málo v trendu vývoje.

Růstové mixture modely, které modelují i variabilitu růstových křivek uvnitř tříd, již byly v popisu variability růstových křivek úspěšnější než výchozí růstové modely, a to i při zohlednění toho, že na to spotřebovaly podstatně více parametrů. Navíc v nich bylo možné dále rozvinout myšlenku hybridních latentních tříd. V modelu se 4 hybridními třídami tak bylo možné uvažovat nad jednou třídou se střední hodnotou autonomie, nevýrazným růstem a velkým reziduálním rozptylem, a třemi třídami s nižšími rezidui – dvěma rostoucími a jednou stagnující. Takový výsledek je pro uvažování nad procesy vývoje autonomie přinejmenším inspirující. Mixture modely vycházející z CFM modelu poskytly výsledky které se od tříd zjištěných na manifestní autonomii mírně liší. V nejjednodušším takovém modelu, kde se třídy nemohly lišit ve variabilitě růstových parametrů uvnitř tříd, se jako nejzajímavější jeví model s 4 třídami. Zakomponování další míry volnosti pro rozdíly mezi latentními třídami již vedlo k problematickým řešení nebo jejich absenci. Výjimkou je opět hybridní model.

S trochou skepse je potřeba konstatovat, že pokud k tomu modely dostaly příležitost, blížily se do značné míry výchozímu latentnímu růstovému modelu s tím, že majoritní třídu doplnili menšími třídami lépe popisujícími část růstových křivek, které majoritní třída již tak dobře nepopisuje. Tyto třídy se však model od modelu liší, a tak je lze považovat spíše za inspirující než za doklad existence skutečných nepozorovaných subpopulací z hlediska vývoje prožívané autonomie.

ZÁVĚRY

Již před více než 10 lety vyjádřil jeden z velkých proponentů růstových mixture modelů Daniel Bauer (2007) skepsi nad potenciálem růstových mixture modelů v psychologii. Jeho výhrady se týkaly toho, že v mnoha dosavadních aplikacích nebyly naplněny některé jejich předpoklady, což vede k nejistotě o validitě extrahovaných latentních tříd. Většina nedodržených předpokladů totiž vede spíše k extrakci příliš mnoha než příliš mála tříd. Odborníka dobře obeznámeného s technickými detaily GMM modelů tak může být obtížně přesvědčit, že extrahované třídy nejsou pouze artefaktem kompenzujícím non-normalitu, nelinearitu či strategii výběr vzorku; a psycholog užívající tyto modely by měl alespoň část této skepse přebírat. To je však do značné míry řešitelné. Technické poznatky o těchto modelech se prudce rozšiřují, takže problémem není ani tak absence řešení neplatnosti toho kterého předpokladu, ale spíše omezená schopnost badatelů se se všemi novými technikami seznámit a využít je. Hlavním zdrojem Bauerovy skepse však v byla nejistota, zda lze na té nejzákladnější úrovni věřit, že v datech skutečně existuje taxonomická struktura, která by mohla být objevena. I když některé GMM analýzy mají předpoklad existence diskrétních, kvalitativně odlišných subpopulací, řada podle Bauera vychází spíše z implicitní motivace založené pouze na tom, že není reálné předpokládat, že by populace byla vývojově homogenní. V tomto smyslu lze pak stěží předpokládat, že by identifikované třídy mohly být skutku *taxonomy* jak je vymezuje například Paul Meehl (1992)

Domnívám se, že užitečnost GMM se vyvíjí a bude vyvíjet stejně jako u dalších komplexních statistických modelů, jichž se psychologie chopila jako heuristických nástrojů – například faktorové analýzy. I přes omezenou statistickou znalost technických detailů modelu se v odborné komunitě vyvinou strategie užití tohoto nástroje, které vedou k plodným teoriím. Matematicko-statistické porozumění modelům, které vede k pochopení toho, jak fungují při aplikaci na data a jak zkeslená mohou být zjištění, je při solidních statistických základech náročné, ale poměrně rychlé. To je však cesta jen mála jednotlivců uvnitř oboru. Většina statistické modely poznává postupně a zkušenostně, zkoumají je podobně jako psychologické fenomény, jimiž se primárně zabývají. Pro lepší porozumění modelu tak potřebují být vystaveni tomu, jak se model chová za různých okolností. Právě v tomto směru si klade tato práce za cíl být užitečná. Předkládá GMM modely a jejich prekurzory v míře detailu, která se do obvyklé výzkumné studie nedostane (byť se dnes množí *online supplements* článků a stoupá množství příznivců transparentní vědy, což činí přístupnými řadu detailů, které dříve přístupně nebývaly).

Prezentování detailů modelovacího procesu facilituje psychologickou interpretaci všech parametrů modelu. Vědomí, že všechny modely jsou odrazem jedněch a těchž dat, je užitečnou brzdou přeinterpretování. Vedle teorie je stabilita výskytu určitého prvku (např. latentní třídy s určitými parametry) napříč více příbuznými modely jedním z užitečných indikátorů toho, že tento prvek není artefaktem modelu. Také usnadňuje „nastavení se“ na metriku závislé proměnné, která je v psychologii typicky velmi arbitrární; obě analýzy v této práci jsou toho příkladem. Teprve růstové modely mě přivedly od primárního interpretování standardizovaných regresních koeficientů k docenění těch nestandardizovaných. Vnímám to jako podobný posun, jako když badatel dokáže přejít od konstatování signifikace, přítomnosti efektu k interpretaci velikosti efektu a souvislosti této velikosti s procesy, které se na něm podílely, ať již z hlediska teorie, či designu a protokolu vzniku dat. Růstové mixture modely jsou komplexní a jejich mnoho parametrů spolu souvisí. Abychom bylo možné tuto

komplexitu v jejím celku dobře interpretovat, je potřeba nejprve dobře rozumět všem parametrům modelu, jejich roli a významu. Bez toho je pravděpodobné, že se analytik zaměří pouze na vrchol pomyslné pyramidy parametrů – tedy na průměry (latentních) růstových koeficientů. Ty však, tím že jsou na vrcholu, jsou ovlivněny parametry, které jsou pod nimi. Proto jsem v popisech modelů v této práci opakovaně explicitně interpretoval parametry modelů.

Podobně jako u EFA se obvykle má za to, že analýza latentních tříd umožňuje „odhalit“ nepozorovanou kategorickou strukturu, která má na svědomí pozorované rozdíly mezi lidmi. A podobně jako u EFA bychom měli být velmi skeptičtí k tomu „odhalování“, protože i růstových mixture modelů máme širokou paletu voleb, které je potřeba při specifikaci modelu učinit, a tyto volby do značné míry ovlivňují výsledek – počet extrahovaných latentních tříd a jejich parametry. Jen zřídka je asi struktura v datech natolik jasná, že se v modelu projeví bez ohledu na detaily specifikace modelu. Proto je při tomto druhu analýzy na místě vyzkoušet více specifikací modelu a sledovat, zda interpretačně konvergují, nebo zda výsledky různých modelů vedou k odlišným představám o kategorické struktuře populace. Jak je zřejmé z předložených analýz, jde o časově náročný postup. Větší jistotu v závěrech pak může poskytnout meta-analýza či systematická přehledová studie více publikovaných mixture analýz. Takových není zatím mnoho, ale nějaké již publikovány byly; například Musliner, Munk-Olsen, Eaton, & Zandi (2016) takto systematizovali studie heterogenity vývoje depresivity.

REZIDUÁLNÍ STRUKTURA JAKO ŽÁDOUCÍ OHNISKO POZORNOSTI

Zaměření této práce na detaily umožnilo věnovat více pozornosti reziduální struktuře růstových i mixture modelů. Ta může být velmi jednoduchá i velmi komplikovaná a v časopisecky publikovaných analýzách se jí obvykle nevěnuje mnoho pozornosti nad rámec prostého popisu. Přitom analýzy prezentované v tomto textu ukazují, že specifikace různých prvků reziduální struktury může model velmi ovlivnit, a dokonce se dostat do interpretačního popředí. V růstových a mixture modelech manifestních proměnných je v každém měření jedno reziduum – rozdíl mezi pozorovanou hodnotou a modelem implikovanou hodnotou v daném čase. Spolu s Grimmem a Widamanem (2010) věřím, že rozptyly reziduí a jejich korelace odrážejí tolik aspektů modelu a modelované charakteristiky, že není jejich specifikaci v modelu nechat na automatických nastavení software. Je tedy dobré zvážit, zda existují důvody věřit, že rezidua by měla být konstantní v čase, nebo ne. Předpoklad homoskedacity je poměrně náročný, protože zahrnuje jak konstantní reliabilitu měření v čase, pravděpodobnostní konstantnost krátkodobých vlivů v čase, i správně zvolený průběh růstové funkce. Možná proto jsou v Mplus automaticky rezidua modelována jako heteroskedastická. Ne snad proto, že by taková nutně měla být, ale aby měl analytik brzy zpětnou vazbu o plauzibilitě či limitech předpokladu homoskedasticity.

O homoskedastická rezidua je důvod usilovat, protože s větším počtem měření znamenají poměrně velkou úsporu parametrů. Na analýze růstových křivek autonomie lze také demonstrovat to, že na základě předpokladu toho, že procesy tvořící modelovanou charakteristiku jsou stabilní, lze fixováním reziduálních rozptylů a dalších parametrů vytvořit model, který spojuje kratší růstové křivky jednotlivců do modelu populace popisující delší období. Tato potřeba nějaké stability může motivovat i poměrně složitě specifikované reziduální struktury, které jsou někde mezi homoskedasticitou a heteroskedasticitou. Příkladem může být reziduální struktura specifikující rezidua tak, aby reliabilita růstové křivky byla

konstantní (Grimm & Widaman, 2010), nebo aby lineárně rostla (McArdle & Nesselroade, 2014).

Rezidua však nejsou pouze otázkou reliability či náhodných faktorů ovlivňujících modelovanou charakteristiku. V psychologii je řada konstruktů, jejichž atributem je nejenom jejich typická úroveň pro daného jedince, ale také nějaká dynamika této úrovně v čase, také typická pro daného jedince. Většina konstruktů související s prožíváním tento atribut má a prožívaná autonomie se nepochybně také patří. Rezidua pak vypovídají i o tom, jak stabilní či labilní pro daného jedince modelovaná proměnná je – jak moc kolísají hodnoty proměnné kolem individuální růstové křivky. Čím větší množství opakovaných měření a dalších dat máme k dispozici, tím stoupá možnost rozlišení mezi tím, zda za rezidui stojí náhodné vnější vlivy, nereliabilita, nebo charakteristická labilita. V tomto smyslu se rezidua mohou v růstové mixture analýze dostat do popřední zájmu, protože jejich vysoká hodnota u části populace může reprezentovat nikoliv náhodu či chybu, ale systematickou charakteristiku. Jak ukazuje práce s rezidui v modelech prožívané autonomie, kombinování latentních růstových koeficientů a reziduí, coby definujících charakteristik latentních tříd, není snadné. Vede k řadě problémů s konvergencí, které jsou patrně způsobeny nedostatečnou empirickou identifikací všech parametrů a velmi zvyšují výpočetní nároky na takovou analýzu. Také může vést, jako v případě autonomie, k nízkým hodnotám entropie. Ty mohou být tak nízké, že zabraňují využít model ke smysluplné kategorizaci, nebo hledání kovariátů členství ve třídě. Podle mého názoru však mají takové modely velkou heuristickou cenu, protože poukazují na vícečetnost procesů, které se podílejí na vývoji modelované charakteristiky.

V souvislosti s množstvím procesů, které spoluutváří modelovaný konstrukt, považuji za vhodné zmínit také korelační vztahy mezi rezidui. Některé modely v tomto textu využívají různé autokorelační struktury, jiné mají rezidua specifikován jako nezávislá. Autokorelační struktury vyjadřují nějakou setrvačnost procesů v čase, která není vysvětlena růstovou křivkou. V případě, kdy zařazení autokorelací zásadním způsobem zlepší shodu modelu s daty, je potřeba se ptát, zda vskutku teoreticky předpokládáme velkou setrvačnost, nebo zda naše pozorování v jednotlivých časech neobsahují informaci o více konstruktech – například proto, že použité měřítko není unidimenzionální, nebo se jeho model měření proměňuje v čase. I zde poskytují růstové modely druhého řádu, které modelují vývoj latentních proměnných spolu s jejich modelem měření, velkou výhodou, protože umožňují předpoklady o modelu měření explicitně ověřovat. Ty oddělují autokorelace disturbancí – tedy křivkou nevysvětleného rozptylu latentních měřítek modelovaného konstruktů – a korelace jednotlivých indikátorů těchto konstruktů. Lze tak rozlišit, co je to, co má vedle růstové křivky také nějakou stabilitu v čase. Je nutné za to však zaplatit daň podobě vyšších nároků na množství dat, velikost vzorku a výpočetní kapacitu.

HYBRIDNÍ SPECIFIKACE LATENTNÍCH TŘÍD

Větší pozornost věnovaná reziduální struktuře mě u modelu autonomie nakonec přivedla k růstovým mixture modelům, v nichž jsou latentní třídy definovány tak, že se mohou lišit mezi sebou v určitých parametrech s výjimkou jedné třídy, která se může lišit ještě v nějakém parametru, který mají ostatní latentní třídy stejný. I když to Mplus umožňuje, zatím jsem se v literatuře s tímto postupem nesešel, a tak jsem jej provizorně pojmenoval *hybridně specifikované latentní třídy*. Takový postup je vlastně jedním krokem od standardních mixture modelů směrem ke konfirmačním mixture modelům (Finch & Bronk, 2011). Přitom při uvažování nad možnými podobami vývoje charakteristik jako je prožívaná autonomie je zcela plauzibilní uvažovat tak, že pro některé subpopulace je definující úroveň modelované

proměnné a její vývoj ve smyslu růstové křivky a pro jiné míra krátkodobé fluktuace hodnot proměnné v čase kolem růstové křivky. Kdyby tyto dva atributy byly nezávislé, kdybychom měli dostatečně velký vzorek, kdyby byly skutečně latentní třídy tako v populaci přítomné a dobře separované a kdybychom měli dostatečnou výpočetní kapacitu (a to je opravdu hodně *kdyby*) bylo by patrně možné umožnit latentním třídám, aby se všechny lišily ve všech parametrech – růstových i reziduálních – a nechat na algoritmu, zda pro takové latentní třídy najde v datech podporu. Reálně je však vhodnější zkombinovat teoretické úvahy s výsledky jednodušších modelů a specifikovat *hybridní latentní třídy*. V případě prožívané autonomie se jak v růstovém mixture modelu, tak v modelu druhého řádu ukázaly jako smysluplné a v souladu s daty modely, kde se jedna třída od ostatních lišila reziduálními rozptyly, tedy fluktuací kolem růstové křivky, a ostatní se mi sebou již lišily pouze průměry a rozptyly latentních růstových koeficientů. Pro subpopulaci s vysoce fluktuující prožívanou autonomií je obtížné určit nějakou růstovou křivku, která by ji dobře vystihovala – ovšem vysoká míra fluktuace implikuje to, že průměrná míra autonomie v této subpopulaci nemůže být příliš vysoká, protože škál autonomie je ovlivněná efektem stropu. Je tedy na místě očekávat, že tato třída se jen stěží může štěpit, kdybychom v modelu specifikovali více tříd. Naopak třídy s nízkým reziduálním rozptylem se od sebe mohou jasněji lišit, a tak lze od modelu očekávat vyšší schopnost je od sebe separovat. V mých analýzách hybridní latentní třídy kombinovaly rozdíly mezi třídami v růstových i reziduálních parametrech a odtud také zdědily svou slabinu, a tou je nízká míra entropie. Mají tak spíše heuristickou cenu než praktickou. Přesto bych je doporučil alespoň ke zvážení při modelování psychologických charakteristik, přinejmenším těch, které mají atribut stability-lability.

VYPOŘÁDÁNÍ SE S CHYBĚJÍCÍMI DATY

Při své komplexnosti jsou růstové modely velmi flexibilní a je v nich možné využít veškerou část dat, kterou máme k dispozici. Analýza prožívané autonomie je příkladem vypořádání se s bohatými, ale komplexně strukturovanými daty. Díky fixování parametrů modelu je možné odhadnout strukturní model i v situaci, kdy některé kovariance ve vstupní kovarianční matici nejsou vůbec odhadnutelné, protože pro ně jednoduše nejsou žádná data. Například mezi pozorovanou autonomií v 18 letech a autonomií v 28 či 27 letech nelze kovarianci odhadnout, protože projekt běžel jen necelých pět let, a tak v datech není nikdo, kdo by měl platnou hodnotu autonomie pro oba věky. Přesto lze za cenu mírného zhoršení ukazatelů shody modelu s daty chybějící prvky kovarianční matice prostřednictvím modelu zafixovat na plauzibilní hodnoty. Je to však možné jen do určité míry, protože při příliš velké míře chybějících dat již přestanou být některé parametry identifikované, což se při rozšíření růstového modelu latentní třídy pochopitelně ještě výrazně zhorší. Zatímco u růstových modelů manifestních hodnot prožívané autonomie bylo možné pracovat s rozpětím od 18 do 28 let, model latentních hodnot prožívané autonomie jsem musel zredukovat na rozpětí 20 až 26 let, protože s přibývajícím komplexitou modelu na cestě od modelu longitudinální invariance přes růstový model až k růstovému mixture modelu bylo odhadnutelné postupně stále menší věkové rozpětí. I když řešení problému s chybějícími daty nebylo ohniskem této práce, přesto se může případný čtenář, který bude čelit podobně komplexním datům, ve specifikacích modelů uvedených v přílohách inspirovat.

LIMITY RŮSTOVÝCH MIXTURE MODELŮ

Vzorek a reprezentativnost. Je přirozené, že růstové mixture modely předpokládají, že data jsou reprezentativním vzorkem populace, jejíž kategorickou strukturu model popisuje. Podle Bauera (2007) je na nereprezentativnost vzorku, respektive na nestejnou pravděpodobnost výběru do vzorku pro jednotlivé členy populace, model zvláště citlivý. Je zřejmé, že pokud jsou členové některé subpopulace výrazně podreprezentováni, popř. nejsou vůbec reprezentováni, nemůže model tuto subpopulaci identifikovat. Pokud je však míra podreprezentace malá, subpopulace ne příliš malá a vzorek velký, identifikována být může. Záleží na tom, jaká data tedy modelu nabídneme, což je opět analogické faktorové analýze, která také nemůže identifikovat faktor, který není faktorovanými proměnnými reprezentován. Taková úvaha může vést ke značné skepsi ohledně možnosti využití růstových mixture modelů v psychologii, protože velké, národně reprezentativní vzorky jsou v našem oboru spíše výjimkou a ani vzorky analyzované v této práci se mezi ně nemohou počítat. Vzorek studie ELSPAC byl sice lokálním cenzem, ale ve věku, z něhož pochází analyzovaná data, již byl značně zredukován procesy, které nelze považovat za náhodné. Vzorek z Cest do dospělosti je pak svou podstatnou spíše příležitostným. I když je zde skepse zcela na místě, domnívám se, že by bylo předčasné užití GMM v těchto konkrétních aplikacích a v psychologii obecně zatracovat. Při vědomí těchto omezení lze výsledky GMM interpretovat. Jen je potřeba reflektovat proces výběru vzorku a vzorce a příčiny úbytku respondentů, uvědomovat si, že model popisuje subpopulace přítomné ve vzorku. To může znamenat, že některé subpopulace resp. latentní třídy mohou chybět, že jejich členové přítomní ve vzorku budou patrně zařazeni do jiných tříd a mohou tak zkreslovat jejich parametry, a že relativní četnosti tříd jsou tím spíše nedohodnocené. Pro utváření teorie to mohou být velmi zajímavé podněty, méně pak pro testování teorie.

TECHNICKÉ A PRAKTICKÉ ZKUŠENOSTI

Většina analýz prezentovaných v této práci byla provedena v programu Mplus v 8. verzi (Muthén & Muthén, 2017). Tento software se vedle Latent GOLD (Vermunt & Magidson 2008) v současné době jeví být jedním z nejjednodušších nástrojů pro provádění komplexních mixture analýz v psychologii. Mezi jeho hlavní přednosti patří rychlost výpočtů, flexibilita a dostupná podpora. Kromě odhadů poskytuje i užitečná zobrazení dat – ta jsou použita u růstových modelů manifestní autonomie. Pro analýzy latentních tříd či mixture analýzy je důležitá možnost dávkového spouštění analýz. Obvykle je potřeba nechat odhadnout sérii modelů s postupně rostoucím počtem latentních tříd. Protože analýzy běží dlouho, není efektivní je spouštět jednu po druhé, ale všechny najednou. Mplus samotný podporuje dávkové spouštění většího počtu .inp souborů, zvláště pro účely Monte Carlo experimentování. I přesto však stojí za to spolu s Mplus využívat vynikající balíček do R prostředí (R Core Team, 2018) **MplusAutomation** (Hallquist & Wiley, 2018). Tento balíček nabízí stále rostoucí paletu funkcí, které umožňují výpočetní schopnosti Mplus používat z prostředí R. V této práci prezentované analýzy rizikového chování a second-order modely autonomie již využívaly funkcí tohoto balíčku. Vedle základní funkce dávkového spouštění většího množství Mplusových modelů, nabízí řadu dalších. Umožňuje dávkově skripty pro Mplus modely vytvářet. Prostřednictvím šablony a několika klíčových slov tak lze vytvořit řadu .inp souborů reprezentující tentýž mixture model s různým počtem latentních tříd, které pak lze dávkově spustit. Další klíčovou funkcí je možnost načíst textové výstupy z Mplus do strukturovaných seznamů (objekty typu *list*) a v R tyto informace dále zpracovávat. Výstupy

mixture analýz mohou být velmi rozsáhlé. Například výstupy z SO-GMM modelů prožívané autonomie měly běžně kolem 5000 řádků a velikost textového souboru se pohybovala kolem půl megabajtu. Již jen uspořádat tyto výstupy za účelem interpretace, natož pak za účelem prezentace, je bez nějaké míry automatizace nesmírně pracné. Tabulky shrnující ukazatele fitu a tabulky s parametry modelu jsem všechny vytvářel s pomocí MplusAutomation.

Jak .inp soubory Mplus, tak R skripty využívající balíček MplusAutomation jsou nedílnou přílohou této práce. Pro sociálně vědného analytika může být logika v těchto skriptech snadnější cestou k porozumění některých aspektů modelování latentních tříd růstových křivek, než jejich matematické či verbálně-interpretativní popisy.

Literatura

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <http://doi.org/10.1109/TAC.1974.1100705>
- Akushevich, I., Kovtun, M., Manton, K. G., & Yashin, A. I. (2009). Linear latent structure analysis and modelling of multiple categorical variables. *Computational and Mathematical Methods in Medicine*, 10(3), 203–218. <http://doi.org/10.1080/17486700802259798>
- Arminger, G., & Stein, P. (1997). Finite Mixtures of Covariance Structure Models with Regressors. *Sociological Methods & Research*, 26(2), 148–182. <http://doi.org/10.1177/0049124197026002002>
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic Latent Class Analysis. *Structural Equation Modeling*, 24(2). <http://doi.org/10.1080/10705511.2016.1253479>
- Asparouhov, T., & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, (14), 1–17. Retrieved from <https://www.statmodel.com/examples/webnotes/webnote14.pdf>
- Asparouhov, T., & Muthén, B. (2014a). Auxiliary Variables in Mixture Modeling: Three-Step Approaches Using M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, (June), 1–13. <http://doi.org/10.1080/10705511.2014.915181>
- Asparouhov, T., & Muthén, B. O. (2014b). Variable-Specific Entropy Contribution. Retrieved from <https://www.statmodel.com/download/UnivariateEntropy.pdf>.
- Asparouhov, T., & Muthén, B. O. (2015). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary second model. Version 2. *Mplus Web Notes*, (21), 1–22.
- Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2016). Relating latent class membership to continuous distal outcomes: improving the LTB approach and a modified three-step implementation. *Structural Equation Modeling*, 23(2), 278–289. <http://doi.org/10.1080/10705511.2015.1049698>
- Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, 23(1), 20–31. <http://doi.org/10.1080/10705511.2014.955104>
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757–786. <http://doi.org/10.1080/00273170701710338>
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3), 338–363. <http://doi.org/10.1037/1082-989X.8.3.338>
- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods*, 9(1), 3–29. <http://doi.org/10.1037/1082-989X.9.1.3>
- Beranová, E., Ježek, S., Širůček, J. (2011). Risk behavior in adolescence: two perspectives. In S. Ježek, L., Lacinová, & P. Macek (Eds.) *Adolescent psychosocial development in Brno: An ELSPAC study 2005–2011*, pp. 107–145. Brno: Masaryk University. ISBN 978-80-210-5682-4.
- Bergman, L. R., Magnusson, D., Khouri, B. M. El, Magnusson, D., & Khouri, B. M. El. (2003). *Studying individual development in an interindividual context*. Psychology Press. <http://doi.org/10.4324/9781410606822>
- Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12(1), 3–27. <http://doi.org/10.1093/pan/mp001>
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370. <http://doi.org>

/10.1007/BF02294361

- Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 372–398. <http://doi.org/10.1080/10705511.2012.687671>
- Cimino, S., Cerniglia, L., Almenara, C. A., Ježek, S., Erriu, M., & Tambelli, R. (2016). Developmental trajectories of body mass index and emotional-behavioral functioning of underweight children: A longitudinal study. *Scientific Reports*, 6(1), 20211. <http://doi.org/10.1038/srep20211>
- Cole, V. T., Bauer, D. J., Hussong, A. M., & Giordano, M. L. (2017). An empirical assessment of the sensitivity of mixture models to changes in measurement. *Structural Equation Modeling*, 24(2). <http://doi.org/10.1080/10705511.2016.1257354>
- Collier, Z. K., & Leite, W. L. (2017). A comparison of three-step approaches for auxiliary variables in latent class and latent profile analysis. *Structural Equation Modeling*, 24(6), 819–830. <http://doi.org/10.1080/10705511.2017.1365304>
- Csémy, L. & Hamanová, J. (2009). Užívání alkoholu českými adolescenty. *Česko-slovenská pediatrie*, 64(10), 454–460.
- De Vos, B. C., Runhaar, J., Verkleij, S. P. J., Van Middelkoop, M., & Bierma-Zeinstra, S. M. A. (2014). Latent class growth analysis successfully identified subgroups of participants during a weight loss intervention trial. *Journal of Clinical Epidemiology*, 67(8). <http://doi.org/10.1016/j.jclinepi.2014.03.007>
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*.
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: frequentist versus Bayesian estimation. *Psychological Methods*, 18(2), 186–219. <http://doi.org/10.1037/a0031609>
- Diallo, T. M. O., & Lu, H. (2017). On the application of the three-step approach to growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 00(00), 1–19. <http://doi.org/10.1080/10705511.2017.1322516>
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. Z. (2016). Impact of misspecifications of the latent variance-covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling*, 23(4), 507–531. <http://doi.org/10.1080/10705511.2016.1169188>
- Donovan, J. E., & Chung, T. (2015). Progressive elaboration and cross-validation of a latent class typology of adolescent alcohol involvement in a national sample. *Journal of Studies on Alcohol and Drugs*, 76(3), 419–429. <http://doi.org/10.15288/jsad.2015.76.419>
- Dowsey, M. M., Smith, A. J., & Choong, P. F. M. (2015). Latent class growth analysis predicts long term pain and function trajectories in total knee arthroplasty: a study of 689 patients. *Osteoarthritis and Cartilage*, 23(12), 2141–2149. <http://doi.org/10.1016/j.joca.2015.07.005>
- Dumenci, L., & Windle, M. (2001). Cluster Analysis as a method of recovering types of intraindividual growth trajectories: a Monte Carlo study. *Multivariate Behavioral Research*, 36(4), 501–522. http://doi.org/10.1207/S15327906MBR3604_02
- Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: concepts, issues, and applications*. (2nd ed.). New York, NY: Lawrence Erlbaum Associates.
- Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling*, 21(4), 534–552. <http://doi.org/10.1080/10705511.2014.919819>
- Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling*, 18(1), 132–151. <http://doi.org/10.1080/10705511.2011.532732>
- Finch, W. H., & French, B. F. (2014). Multilevel latent class analysis: parametric and nonparametric models. *Journal of Experimental Education*, 82(3), 307–333. <http://doi.org/>

- 10.1080/00220973.2013.813361
- Fonseca-Pedrero, E., Wells, C., Paino, M., Lemos-Giráldez, S., Villazón-García, Ú., Sierra, S., ... Muñiz, J. (2010). Measurement invariance of the reynolds depression adolescent scale across gender and age. *International Journal of Testing*, *10*(2), 133–148. <http://doi.org/10.1080/15305050903580822>
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time . . . or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*. <http://doi.org/10.1037/pas0000275>
- Geiser, C., Keller, B. T., & Lockhart, G. (2013). First- versus second-order latent growth curve models: some insights from latent state-trait theory. *Structural Equation Modeling*, *20*(3), 479–503. <http://doi.org/10.1080/10705511.2013.797832>
- Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state–trait analyses. *Psychological Methods*, *17*(2), 255–283. <http://doi.org/10.1037/a0026977>
- Grilli, L., & Varriale, R. (2014). Specifying measurement error correlations in latent growth curve models with multiple indicators. *Methodology*, *10*(4), 117–125. <http://doi.org/10.1027/1614-2241/a000082>
- Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in M plus and SAS. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 676–701. <http://doi.org/10.1080/10705510903206055>
- Grimm, K. J., & Widaman, K. F. (2010). Residual Structures in latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 424–442. <http://doi.org/10.1080/10705511.2010.489006>
- Gruber, E., DiClemente, R. J., Anderson, M. M., & Lodico, M. (1996). Early drinking onset and its association with alcohol use and problem behavior in late adolescence. *Preventive Medicine*, *25*(3), 293–300. <http://doi.org/10.1006/PMED.1996.0059>
- Guerra-Peñ, K., Steinley, D., & Guerra-Peña, K. (2016). Extracting spurious latent classes in growth mixture modeling with nonnormal errors growth mixture modeling with normal errors assumption. *Educational and Psychological Measurement*, *76*(6), 933–953. <http://doi.org/10.1177/0013164416633735>
- Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling*, *25*(4), 621–638. <http://doi.org/10.1080/10705511.2017.1402334>
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, New Jersey: John Wiley & Sons.
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*(2), 202–226. <http://doi.org/10.1080/10705510709336744>
- Hertzog, C., Ghisletta, P., Lindenberger, U., & von Oertzen, T. (2006). On the Power of Multivariate Latent Growth Curve Models to Detect Correlated Change. *Psychological Methods*, *11*(3), 244–252. <http://doi.org/10.1037/1082-989X.11.3.244>
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36–53. <http://doi.org/10.1037/1082-989X.11.1.36>
- IBM. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.
- Isiordia, M., & Ferrer, E. (2018). Curve of Factors Model: A Latent Growth Modeling Approach for Educational Research. *Educational and Psychological Measurement*, *78*(2), 203–231. <http://doi.org/10.1177/0013164416677143>
- Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2017). A comparison of methods for uncovering sample heterogeneity: structural equation model trees and finite mixture models. *Structural Equation Modeling*, *24*(2), 270–282. <http://doi.org/10.1080/>

10705511.2016.1250637

- Ježek, S. (2014). Aktuální pojetí autonomie v psychologii. *Československá psychologie, Academia*, roč. 58, č. 1, s. 31-40. ISSN 0009-062X.
- Ježek, S., Lacinová, L. & Macek, P. (Eds.) (2011). *Adolescent psychosocial development in Brno: An ELSPAC study 2005 – 2011*. Brno: Masaryk University. ISBN 978-80-210-5682-4.
- Ježek, S., Macek, P. & Bouša, O. (2016). Cesty k nezávislosti: Jak se vyvíjí autonomie s identitou. In L. Lacinová, S. Ježek, P. Macek. *Cesty do dospělosti: Psychologické a sociální charakteristiky dnešních dvacátníků*. 1. vyd. Brno: MUNI Press. s. 25-38, 14 s. Psychologie. ISBN 978-80-210-8400-1.
- Ježek, S., Masopustová, Z. & Bouša, O. (2011). *Vývoj depresivity v průběhu adolescence a moderátory růstových křivek*. Příspěvek prezentovaný na konferenci Psychologická diagnostika dětí a dospívajících: Výzkum, prevence a školní poradenství, June 29 – July 1, Brno.
- Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302–317. <http://doi.org/10.1111/j.1751-9004.2007.00054.x>
- Kenny, D. A. (2018). Dostupné online: <http://davidakenny.net/cm/long.htm>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. <http://doi.org/10.1177/0049124114543236>
- Kreuter, F., & Muthén, B. O. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology* 24. <http://doi.org/10.1007/s10940-007-9036-0>
- Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: a flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1), 1–26. <http://doi.org/10.1080/10705511.2013.742377>
- Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, 46(2), 266–302. <http://doi.org/10.1080/00273171.2011.556549>
- Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, 74(4), 557–584. <http://doi.org/10.1177/0013164413519798>
- Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. <http://doi.org/10.1037/met0000075>
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778.
- Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21–39. <http://doi.org/10.1037/1082-989X.10.1.21>
- Luoma, I., Korhonen, M., Salmelin, R. K., Helminen, M., & Tamminen, T. (2015). Long-term trajectories of maternal depressive symptoms and their antenatal predictors. *Journal of Affective Disorders*, 170, 30–38. <http://doi.org/10.1016/j.jad.2014.08.017>
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2), 191–225. <http://doi.org/10.1080/10705510902751010>
- Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling*, 25(3), 389–402. <http://doi.org/10.1080/>

10705511.2017.1389611

- McArdle, J. J. (1988). Dynamic but Structural Equation Modeling of Repeated Measures Data. In J. R. Nesselroade & R. B. Cattell R. B. (Eds.), *Handbook of Multivariate Experimental Psychology. Perspectives on Individual Differences*, pp. 561–614. Springer: Boston, MA
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- McLarnon, M. J. W., & O'Neill, T. A. (2018). Extensions of auxiliary variable approaches for the investigation of mediation, moderation, and conditional effects in mixture models. *Organizational Research Methods*, 21(4), 955–982. <http://doi.org/10.1177/1094428118770731>
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, 60(1), 117–173.
- Meredith, W. & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In Collins, L. M. & Sayer, A. G. (Eds.), *New methods for the analysis of change*, pp. 203–240; Washington, DC, US: American Psychological Association;
- Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: an introduction. *Child Development Perspectives*, 4(1), 5–9. <http://doi.org/10.1111/j.1750-8606.2009.00109.x>
- Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders*, 192, 199 - 211. <http://doi.org/10.1016/j.jad.2015.12.030>
- Muthén, B., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 639–657. <http://doi.org/10.1111/j.1467-985X.2009.00589.x>
- Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: APA. Retrieved from [papers3://publication/uuid/D2E64C4E-7C99-45CB-9377-3966A16F5398](https://pubs3://publication/uuid/D2E64C4E-7C99-45CB-9377-3966A16F5398)
- Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, 8, 369–377.
- Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.
- Muthén, B.O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis*, pp. 143–165. Boca Raton: Chapman & Hall/CRC.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2(4), 371–402. <http://doi.org/10.1037/1082-989X.2.4.371>
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research*, 24(6), 882–91. <http://doi.org/10.1111/j.1530-0277.2000.tb02070.x>
- Muthén, B. O., & Muthén, L. K. (2017). *Mplus user's guide* (8th ed). Los Angeles, CA: Muthén & Muthén.
- Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2), 463–469. Retrieved from <http://www.jstor.org/stable/2533793>
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2), 139–157. <http://doi.org/10.1037/1082-989X.4.2.139>

- Nagin, D. S. (2009). *Group-based modeling of development*. Harvard University Press.
- Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling, 23*(6). <http://doi.org/10.1080/10705511.2016.1221313>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569. <http://doi.org/10.1080/10705510701575396>
- Osecká, L. (1999). *Aplikace shlukové analýzy v psychologii osobnosti*. Habilitační práce, Brno: Masarykova univerzita.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Raftery, A.E. (1995) Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995*, pp 111-164. Oxford: Blackwell.
- Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: a method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioural Development, 33*(6), 565–576. <http://doi.org/10.1177/0165025409343765>.
- Raudenbush, S. W. (2005). How do we study “what happens next”? *Annals of the American Academy of Political and Social Science, 602*, 131–144.
- Robinson, K. A., Perez, T., Nuttall, A. K., Roseth, C. J., & Linnenbrink-Garcia, L. (2018). From science student to scientist: Predictors and outcomes of heterogeneous science identity trajectories in college. *Developmental Psychology, 54*(10), 1977–1992. <http://doi.org/10.1037/dev0000567>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.
- Slove, L. S. (1987). Application of model selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333–343.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. (1st ed.). New York, NY: Oxford University Press.
- Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). Lasagna plots: A saucy alternative to spaghetti plots. *Epidemiology, 21*(5), 621–625. <http://doi.org/10.1097/EDE.0b013e3181e5b06a>
- Širůček, J. & Širůčková, M. (2008). Risk behavior. In S. Ježek, L., Lacinová (Eds.), *Fifteen-Year-Olds in Brno: A Slice of Longitudinal Self-Reports*, pp. 25 – 38. Brno: Masaryk University. ISBN 978-80-210-4755-6.
- Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 640–657. <http://doi.org/10.1080/10705511.2013.824781>
- Tekle, F. B., Gudicha, D. W., & Vermunt, J. K. (2014). Power analysis for the Bootstrap Likelihood Ratio Test in Latent Class Models, 1–20. Retrieved from <https://pdfs.semanticscholar.org/7a8c/2275a01b705bf51a6bda2bda33d9789f6623.pdf>
- Tekle, F. B., Gudicha, D. W., & Vermunt, J. K. (2016). Power analysis for the bootstrap likelihood ratio test for the number of classes in latent class models. *Advances in Data Analysis and Classification, 10*(2), 209–224. <http://doi.org/10.1007/s11634-016-0251-0>
- Tofighi, D., & Enders, C. (2006). Identifying the correct number of classes in growth mixture models. In G. Hancock & K. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC.
- Tueller, S., & Lubke, G. (2010). Evaluation of Structural Equation Mixture Models: Parameter Estimates and Correct Class Assignment. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(2), 165–192. <http://doi.org/10.1080/10705511003659318>
- van den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2017). Building latent class trees,

- with an application to a study of social capital. *Methodology*, 13(Supplement 1), 13–22. <http://doi.org/10.1027/1614-2241/a000128>
- Van Horn, M. L., Feng, Y., Kim, M., Lamont, A., Feaster, D., & Jaki, T. (2016). Using multilevel regression mixture models to identify level-1 heterogeneity in level-2 effects. *Structural Equation Modeling*, 23(2). <http://doi.org/10.1080/10705511.2015.1035437>
- Wang C.P., Brown, C.H., Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, 100 (3), 1054-1076.
- Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17(1), 33–51. <http://doi.org/10.1177/0962280207081238>
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4), 450–469. <http://doi.org/10.1093/pan/mpq025>
- Vermunt, J. K., & Dijk, L. A. Van. (2001). A non-parametric random-coefficient approach: The latent class regression model. *Multilevel Modelling Newsletter*, 13(2), 6–13.
- Vermunt J. K., Magidson, J. (2008). *Manual for latent GOLD 4.5 syntax module*. Statistical Innovations Inc, Belmont, MA.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2), 307. <http://doi.org/10.2307/1912557>
- Wickrama, K. K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher-Order Growth Curves and Mixture Modeling with Mplus: A Practical Guide*. New York, NY: Routledge.
- Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics and Data Analysis*, 50(4), 1090–1104. <http://doi.org/10.1016/j.csda.2004.11.004>
- Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62(3), 297–330. <http://doi.org/10.1007/BF02294554>

Příloha 1 – Specifikace prezentovaných modelů v Mplus

RIZIKOVÉ CHOVÁNÍ

LGCM

```
TITLE: Unconditional linear LGGM for Risk Behavior Scale 2 - substance abuse
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = id n pohlavi s1_13 s1_15 s1_17 s1_19 s2ln_13 s2ln_15 s2ln_17 s2ln_19
                 s3ln_13 s3ln_15 s3ln_17 s3ln_19 s3dich_13 s3dich_15 s3dich_17 s3dich_19
                 s4i_13 s4i_15 s4i_17 s4i_19 s5ln_13 s5ln_15 s5ln_17 s5ln_19
                 Zimpulsiv cap_pos cap_neg cap_aut ZPA_S ZPA_R ZPA_D
                 Zinform Znformat Zdifffus filter;
MISSING = ALL (999);
USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19;
USEOBSERVATIONS = n > 1; ! Omezení pouze na ty, kdo mají alespoň 2 měření

ANALYSIS:
  ESTIMATOR = MLR;
  INFORMATION = OBSERVED;

MODEL:
  I S | s2ln_13@0 s2ln_15@2 s2ln_17@4 s2ln_19@6; !Lineární růst
  I S; !Rozptyly latentního průsečíku a směrnice
  I with S; !Kovariance latentního průsečíku a směrnice
  s2ln_13 s2ln_15 s2ln_17 s2ln_19; !Reziduální rozptyly

OUTPUT: SAMPSTAT STDYX PATTERNS TECH7;

PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);
```

Varianta specifikace lineárního modelu s homoskedastickými rezidui

```
MODEL:
  I S | s2ln_13@0 s2ln_15@2 s2ln_17@4 s2ln_19@6;
  I S;
  I with S;
  s2ln_13-s2ln_19 (r); !Všechny reziduální rozptyly stejné.
```

Varianta specifikace lineárního modelu s autokorelační strukturou 1. řádu

```
MODEL:
  I S | s2ln_13@0 s2ln_15@2 s2ln_17@4 s2ln_19@6;
  I S;
  I with S;
  s2ln_13 s2ln_15 s2ln_17 s2ln_19;
  s2ln_13-s2ln_17 PWITH s2ln_15-s2ln_19; !Autokorelace reziduálních rozptylů
```

Specifikace latent-base modelu

```
MODEL:
  I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6; !Uvolnění časových bází v 15 a 17 letech
  I S;
  I with S;
  s2ln_13 s2ln_15 s2ln_17 s2ln_19;
```

Specifikace modelu s logistickým růstem

```
MODEL:
  I by s2ln_13 - s2ln_19@1; !Alternativní specifikace interceptu pomocí BY
  S by s2ln_13* (L1)
```

```

s2ln_15 (L2)
s2ln_17 (L3)
s2ln_19 (L4);      !Časové báze pojmenovány L1-L4 a odhadovány podle MODEL
CONSTRAINT
I S;
I with S;
[I-S];              !Průměry latentního průsečíku a směrnice
s2ln_13-s2ln_19 (e);      !Reziduální rozptyly
[s2ln_13-s2ln_19@0];      !Průsečíky manifestních indikátorů

MODEL CONSTRAINT:
NEW(alpha*.5 lambda*2);  !Parametry logistické růstové křivky a jejich startovací hodnoty
L1 = 1/(1 + EXP(-(0-lambda)*alpha));  !Časová báze je odvozena z alfy a lambda
L2 = 1/(1 + EXP(-(2-lambda)*alpha));
L3 = 1/(1 + EXP(-(4-lambda)*alpha));
L4 = 1/(1 + EXP(-(6-lambda)*alpha));

```

Specifikace modelu s Richardsovým růstem

```

MODEL:
I by s2ln_13 - s2ln_19@1; !Alternativní specifikace interceptu pomocí BY
S by s2ln_13* (L1)
    s2ln_15 (L2)
    s2ln_17 (L3)
    s2ln_19 (L4); !Časové báze pojmenovány L1-L4 a odhadovány podle MODEL CONSTRAINT
I S;
I with S;
[I-S];              !Průměry latentního průsečíku a směrnice
s2ln_13-s2ln_19 (e);      !Reziduální rozptyly
[s2ln_13-s2ln_19@0];      !Průsečíky manifestních indikátorů

MODEL CONSTRAINT:
NEW(alpha*.5 lambda*2 tau*.5); !Parametry Richardsovy křivky a jejich startovací hodnoty
L1 = 1/ ( (1 + tau*EXP(-(0-lambda)*alpha))^(1/tau) );
L2 = 1/ ( (1 + tau*EXP(-(2-lambda)*alpha))^(1/tau) );
L3 = 1/ ( (1 + tau*EXP(-(4-lambda)*alpha))^(1/tau) );
L4 = 1/ ( (1 + tau*EXP(-(6-lambda)*alpha))^(1/tau) );

```

LCGA

LCGA latent-base model se třemi třídami lišícími se v průměrech I a S.

```

TITLE: Unconditional latent base LGGA for Risk Behavior Scale 2 - substance abuse
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
    MISSING = ALL (999);
    USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19;
    USEOBSERVATIONS = n > 1;
    CLASSES = C(3);

ANALYSIS:
    TYPE=MIXTURE;
    ESTIMATOR = MLR;

!Sekce pro hlavní analýzu
!STARTS = 800 160;      !Počet náhodných sad startovacích hodnot a finálních optimalizací

!Sekce pro LMR test (TECH11) pro BLRT test (TECH14)
STARTS = 0;
OPTSEED=263049;        !Seed, s nímž bylo dosaženo v hlavní analýze nejnižší -LL
K-1STARTS = 800 160;  !Startovacích hodnoty a finálních opt. pro model s k-1 třídami
LMR
    !LRTBOOTSTRAP = 500; !Bez specifikace počtu BS vzorků, Mplus použije optional stopping

```

```

LRTSTARTS = 20 4 800 160;  !Startovacích hodnoty a finálních opt.
                             !pro model s k-1 třídami a k třídami - BLRT

MODEL:
%OVERALL%  !Protože se třídy liší jen průměry I a S, stačí specifikovat jen celkový
model
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;          !Růstová křivka
I@0;                                                  !Rozptyl průsečíků i směrníc
fixován na 0
S@0;
I with S@0;                                          !... a tedy i jejich kovariance
s2ln_13 s2ln_15 s2ln_17 s2ln_19;

OUTPUT: SAMPSTAT TECH7;
        TECH11;      !LMR test - vypnout v hlavní analýze.
        TECH14;      !BLRT test - vypnout v hlavní analýze.

PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);

```

Specifikace LCGA latent-base modelu se třemi třídami lišícími se v průměrech I a S a tvarech růstových křivek.

```

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I@0;
S@0;
I with S@0;
s2ln_13 s2ln_15 s2ln_17 s2ln_19;

%c#2%          !Sekce pro parametry, které se mají lišit ve třídě 2
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;          !Odhad časových bází pro třídu 2

%c#3%          !Sekce pro parametry, které se mají lišit ve třídě 3
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;          !Odhad časových bází pro třídu 3

```

GMM

GMM latent-base model se třemi třídami lišícími se v průměrech a rozptylech I a S.

```

TITLE: Unconditional latent-base GMM for Risk Behavior Scale 2 - substance abuse
       Growth factor variance varying across classes.
       Latent-bases constant across classes.
       Rezidual variance fixed.

DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);
          USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19;
          USEOBSERVATIONS = n > 1;
          CLASSES = C(3);

ANALYSIS:
PROCESSORS = 12;
TYPE = MIXTURE;
ESTIMATOR = MLR;

!Sekce pro hlavní analýzu
STARTS = 3200 320;

```

```

!Sekce pro LMR test (TECH11) pro BLRT test (TECH14)
!STARTS = 0;
!OPTSEED=496703;
!K-1STARTS = 80 16;
!LRTBOOTSTRAP = 10;
!LRTSTARTS = 20 4 800 160;

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);

%c#2%
I-S;
I with S;

%c#3%
I-S;
I with S;

OUTPUT: SAMPSTAT TECH7;
        !TECH11;      !LMR test.
        !TECH14;      !BLRT test.

PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);

```

Specifikace GMM latent-base modelu se třemi třídami lišícími se v průměrech a rozptylech I a S i ve tvaru růstové křivky.

```

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;

%c#3%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;

```

GMM modely s kovariáty

Pohlaví jako spojitý prediktor členství v latentní třídě. Jednokrokový přístup.

```

TITLE: GMM for Risk Behavior Scale 2 - substance abuse with sex as a covariate
      Growth factors variance varying across classes. Latent-base varying across
classes.
      Rezidual variance fixed.
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);
          USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19 sex;

```



```

        USEOBSERVATIONS = n > 1;
        IDVARIABLE = id;
        CLASSES = C(2);

ANALYSIS:
PROCESSORS = 12;
TYPE = MIXTURE;
ESTIMATOR = MLR;
STARTS = 800 160;

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);

C ON sex;

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;

OUTPUT: SAMPSTAT STDYX REZIDUAL TECH7;

PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);

```

Pohlaví jako spojitý prediktor členství v latentní třídě. Tříkrokový přístup. Krok 1 – export členství ve třídě

```

TITLE: ...
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);
          USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19;
          USEOBSERVATIONS = n > 1;
          IDVARIABLE = id;           !Identifikační proměnná pro slučování dat
          CLASSES = C(2);

ANALYSIS:
TYPE = mixture;
ESTIMATOR = MLR;

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;

OUTPUT: SAMPSTAT TECH7;
        ENTROPY;
PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);

SAVEDATA: FILE = "S2_GM2_2_sex6.dat";   ! Export
      SAVE = CPROB;

```

Krok 2 – dummy model

```
TITLE: ...
DATA: FILE="..\..\S2_GM2_2_sex6a.dat";      !Data s připojenými proměnnými o členství
VARIABLE: NAMES = ...
          cprob1 cprob2 class;
          NOMINAL = class;      !Stanovíme nejpravděpodobnější třídu jako nominální.
          MISSING = ALL (999);
          USEVAR = class;
          IDVARIABLE = id;
          CLASSES = C(2);

ANALYSIS:
TYPE = MIXTURE;
ESTIMATOR = MLR;

MODEL:
%OVERALL%

%c#1%
[class#1@4.005];

%c#2%
[class#1@-0.506];

OUTPUT: SAMPSTAT; TECH7;
```

Krok 3 – dummy model s kovariátem

```
MODEL:
%OVERALL%
C ON sex;

%c#1%
[class#1@4.005];

%c#2%
[class#1@-0.506];
```

Pohlaví jako spojitý prediktor členství v latentní třídě. R3STEP (BCH, DCON) technika.

```
TITLE: GMM for Risk Behavior Scale 2 - substance abuse with sex as a covariate
      Growth factors variance varying across classes. Latent-base varying across
classes.
      Residual variance fixed.
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);
          USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19 sex;
          USEOBSERVATIONS = n > 1;
          IDVARIABLE = id;
          CLASSES = C(2);
          AUXILIARY = sex (R3STEP);      !Stejně lze použít i BCH či DCON
...
...

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
```

```

I-S;
I with S;
s2ln_13-s2ln_19 (e);

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
...

```

Pohlaví jako prediktor I a S uvnitř každé třídy zvlášť. Jednokroková technika.

TITLE: GMM for Risk Behavior Scale 2 - substance abuse with sex as a covariate
 Growth factors variance varying across classes. Latent-base varying across
 classes.

```

Rezidual variance fixed.
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);
          USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19 sex;
          USEOBSERVATIONS = n > 1;
          IDVARIABLE = id;
          CLASSES = C(2);
...

```

```

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);
I-S ON sex;

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
I-S ON sex;
...

```

Pohlaví jako prediktor I a S uvnitř každé třídy zvlášť i jako prediktor členství ve třídě. Jednokroková technika.

```

MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);

C ON sex;
I-S ON sex;

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
I-S ON sex;

```

Pohlaví jako kategorická proměnná závislá na členství ve třídě

```
VARIABLE: ...
      CATEGORICAL = sex;
MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);
[sex$1] (T1);

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
[sex$1] (T2);

MODEL TEST:
T1=T2;
```

Pohlaví jako spojitá proměnná závislá na členství ve třídě

```
MODEL:
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);
[sex] (M1);
sex;

%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
[sex] (M2);
sex;

MODEL TEST:
M1=M2;
```

PROŽÍVANÁ AUTONOMIE

Modely prožívané autonomie jsou analogické modelům rizikového chování. Uvádím zde modely, které jsou něčím specifické.

Nulový LGC model

Korektní specifikace nulového modelu pro výpočet CFI a TLI.

```
TITLE: Unconditional mean LGCM for autonomy - null model for CFI
DATA: FILE="Awide.dat";
VARIABLE: NAMES = id sex pocaut aut18 autsq18 mood18 sds18 aut19 autsq19 mood19 sds19
      aut20 autsq20 mood20 sds20 aut21 autsq21 mood21 sds21 aut22 autsq22 mood22 sds22
      aut23 autsq23 mood23 sds23 aut24 autsq24 mood24 sds24 aut25 autsq25 mood25 sds25
      aut26 autsq26 mood26 sds26 aut27 autsq27 mood27 sds27 aut28 autsq28 mood28 sds28;
      MISSING = ALL (999);
      USEVAR = autsq18 autsq19 autsq20 autsq21 autsq22
      autsq23 autsq24 autsq25 autsq26 autsq27 autsq28;
      USEOBSERVATIONS = pocaut > 2;      !Omezení pouze na ty, kdo vydrželi 3 roky
```

a více.

ANALYSIS: COVERAGE=0;

!Není nikdo, kdo by měl naměřeno v 18 i 28.

MODEL:

I | autsq18@0 autsq19@1 autsq20@2 autsq21@3 autsq22@4 autsq23@5
autsq24@6 autsq25@7 autsq26@8 autsq27@9 autsq28@10;
I@0;

OUTPUT: SAMPSTAT PATTERNS TECH1 TECH7;

PLOT: TYPE=PLOT3;

SERIES = autsq18-autsq28(*);

LGC model s konstantní reliabilitou růstové křivky dle (Grimm, Widaman, 2010)

MODEL:

I by autsq18@1
autsq19@1
autsq20@1
autsq21@1
autsq22@1
autsq23@1
autsq24@1
autsq25@1
autsq26@1
autsq27@1
autsq28@1;

I (v0);

[I];

S by autsq18@0
autsq19@1
autsq20@2
autsq21@3
autsq22@4
autsq23@5
autsq24@6
autsq25@7
autsq26@8
autsq27@9
autsq28@10;

S (v1);

[S];

I with S (c01);

autsq18 (ve0);
autsq19 (ve1);
autsq20 (ve2);
autsq21 (ve3);
autsq22 (ve4);
autsq23 (ve5);
autsq24 (ve6);
autsq25 (ve7);
autsq26 (ve8);
autsq27 (ve9);
autsq28 (ve10);

[autsq18@0];
[autsq19@0];
[autsq20@0];
[autsq21@0];
[autsq22@0];
[autsq23@0];
[autsq24@0];
[autsq25@0];
[autsq26@0];

```

[autsq27@0];
[autsq28@0];

MODEL CONSTRAINT:
NEW (rho*.8);
ve0 = ((v0) - (v0 * rho))/rho;
ve1 = ((V0+1*V1*1+2*(c01*1))-((V0+1*V1*1+2*(c01*1))*rho))/rho;
ve2 = ((V0+2*V1*2+2*(c01*2))-((V0+2*V1*2+2*(c01*2))*rho))/rho;
ve3 = ((V0+3*V1*3+2*(c01*3))-((V0+3*V1*3+2*(c01*3))*rho))/rho;
ve4 = ((V0+4*V1*4+2*(c01*4))-((V0+4*V1*4+2*(c01*4))*rho))/rho;
ve5 = ((V0+5*V1*5+2*(c01*5))-((V0+5*V1*5+2*(c01*5))*rho))/rho;
ve6 = ((V0+6*V1*6+2*(c01*6))-((V0+6*V1*6+2*(c01*6))*rho))/rho;
ve7 = ((V0+7*V1*7+2*(c01*7))-((V0+7*V1*7+2*(c01*7))*rho))/rho;
ve8 = ((V0+8*V1*8+2*(c01*8))-((V0+8*V1*8+2*(c01*8))*rho))/rho;
ve9 = ((V0+9*V1*9+2*(c01*9))-((V0+9*V1*9+2*(c01*9))*rho))/rho;
ve10 = ((V0+10*V1*10+2*(c01*10))-((V0+10*V1*10+2*(c01*10))*rho))/rho;

```

Piecewise model

```

MODEL:
I S1 | autsq18@0 autsq19@1 autsq20@2 autsq21@3 autsq22@4 autsq23@5
      autsq24@6 autsq25@7 autsq26@8 autsq27@8 autsq28@8;
I S2 | autsq18@0 autsq19@0 autsq20@0 autsq21@0 autsq22@0 autsq23@0
      autsq24@0 autsq25@0 autsq26@1 autsq27@2 autsq28@3;

```

Specifikace „hybridního“ GMM modelu se třemi třídami, s jednou reziduálně odlišnou třídou

```

MODEL:
%OVERALL%
eta0 etal | autsq18@0 autsq19@1 autsq20@2 autsq21@3 autsq22@4 autsq23@5
           autsq24@6 autsq25@7 autsq26@8 autsq27@9 autsq28@10;

%c#1%
eta0-etal ;
[eta0-etal];
eta0 with etal;
autsq18-autsq28 (e1); !Rezidua v této třídě jsou fixována na e1

%c#2%
eta0-etal ;
[eta0-etal];
eta0 with etal;
autsq18-autsq28 (e2); !Rezidua v této a následujících třídách jsou fixována na e2

%c#3%
eta0-etal;
[eta0-etal];
eta0 with etal;
autsq18-autsq28 (e2);

```

Specifikace „hybridního“ GMM modelu se třemi třídami, s jednou odlišnou třídou

Syntax je z fáze, kdy již se odhadovaly LRT testy. Proto jsou STARTS na 0 a startovací parametry vygenerovány podle OPTSEED z již proběhnuvšího modelu.

```

TITLE: Unconditional linear GMM for autonomy
DATA: FILE="Awide.dat";
VARIABLE: NAMES = ...
          MISSING = ALL (999);

```

```

USEVAR = autsq18 autsq19 autsq20 autsq21 autsq22
        autsq23 autsq24 autsq25 autsq26 autsq27 autsq28;
USEOBSERVATIONS = pocaut > 2;    ! Omezení pouze na ty, kdo vydrželi N roků.
CLASSES=C(3);

ANALYSIS:
TYPE = MIXTURE;
PROCESSORS = 12;
COVERAGE=0;          !Nejsou ti, co by měli naměřeno v 18 i 28.

!Sekce pro hlavní analýzu
!STARTS = 400 80;          !Default is 20 4

!Sekce pro LMR test (TECH11) a BLRT test (TECH14)
OPTSEED=605358;
STARTS = 0;
K-1STARTS = 400 80;
LRTBOOTSTRAP = 300;
LRTSTARTS = 0 0 300 60;

MODEL:
%OVERALL%
eta0 etal | autsq18@0 autsq19@1 autsq20@2 autsq21@3 autsq22@4 autsq23@5
           autsq24@6 autsq25@7 autsq26@8 autsq27@9 autsq28@10;

%c#1%
eta0-etal ;
[eta0-etal];
eta0 with etal;
autsq18-autsq28 (e1);

%c#2%
eta0-etal (v1-v2);          ! Rozptyly, které mají být od 2. třídy stejné
[eta0-etal];
eta0 with etal (c1);       ! Kovariance, která má být od 2. třídy stejná
autsq18-autsq28 (e2);

%c#3%
eta0-etal (v1-v2);
[eta0-etal];
eta0 with etal (c1);
autsq18-autsq28 (e2);

OUTPUT: SAMPSTAT STDYX TECH7;
        TECH11;          !LMR test.
        TECH14;          !BLRT test.

```

CFM model autonomie s logistickým růstem

```

TITLE: Unconditional Curve of Factors Model for autonomy.
       CFA model with items 5,6 allowed to have correlated residuals.
       Věky 20 - 26.
       Logistický růst s lag1 covariancemi a konstantními disturbancemi.
DATA: FILE="AIwide3.dat";
VARIABLE: NAMES = id sex zacatek konec delka pocaut
           A1_18 A2_18 A3_18 A4_18 A5_18 A6_18 A7_18 mood_18 sds_18
           A1_19 A2_19 A3_19 A4_19 A5_19 A6_19 A7_19 mood_19 sds_19
           A1_20 A2_20 A3_20 A4_20 A5_20 A6_20 A7_20 mood_20 sds_20
           A1_21 A2_21 A3_21 A4_21 A5_21 A6_21 A7_21 mood_21 sds_21
           A1_22 A2_22 A3_22 A4_22 A5_22 A6_22 A7_22 mood_22 sds_22
           A1_23 A2_23 A3_23 A4_23 A5_23 A6_23 A7_23 mood_23 sds_23
           A1_24 A2_24 A3_24 A4_24 A5_24 A6_24 A7_24 mood_24 sds_24
           A1_25 A2_25 A3_25 A4_25 A5_25 A6_25 A7_25 mood_25 sds_25

```

```

A1_26 A2_26 A3_26 A4_26 A5_26 A6_26 A7_26 mood_26 sds_26
A1_27 A2_27 A3_27 A4_27 A5_27 A6_27 A7_27 mood_27 sds_27
A1_28 A2_28 A3_28 A4_28 A5_28 A6_28 A7_28 mood_28 sds_28;
MISSING = ALL (999);
USEVAR = A1_20-A7_20 A1_21-A7_21 A1_22-A7_22 A1_23-A7_23
        A1_24-A7_24 A1_25-A7_25 A1_26-A7_26;
USEOBS = (konec > 20) AND (zacatek < 26) AND (pocaut > 2);

ANALYSIS: ESTIMATOR=ML;
          COVERAGE=0;
          H1ITERATIONS = 50000;
          H1CONVERGENCE = 0.001;
          PROCESSORS = 12;

MODEL:
!Náboje
  A20 BY A1_20-A7_20 (L1-L7); !Fixovány
  A21 BY A1_21-A7_21 (L1-L7);
  A22 BY A1_22-A7_22 (L1-L7);
  A23 BY A1_23-A7_23 (L1-L7);
  A24 BY A1_24-A7_24 (L1-L7);
  A25 BY A1_25-A7_25 (L1-L7);
  A26 BY A1_26-A7_26 (L1-L7);
!Průsečíky položek
  [A1_20@0]; !Fixování průsečíku marker proměnné na 0...
  [A1_21@0]; !...pro identifikaci průměrů faktorů.
  [A1_22@0];
  [A1_23@0];
  [A1_24@0];
  [A1_25@0];
  [A1_26@0];
  [A2_20-A7_20] (I2-I7); !Fixovány
  [A2_21-A7_21] (I2-I7);
  [A2_22-A7_22] (I2-I7);
  [A2_23-A7_23] (I2-I7);
  [A2_24-A7_24] (I2-I7);
  [A2_25-A7_25] (I2-I7);
  [A2_26-A7_26] (I2-I7);
!Průměry faktorů
  [A20-A26@0]; !Pro identifikaci průměrů I a S.

!Kovariance reziduí v rámci jednoho měření
  A5_20 with A6_20;
  A5_21 with A6_21;
  A5_22 with A6_22;
  A5_23 with A6_23;
  A5_24 with A6_24;
  A5_25 with A6_25;
  A5_26 with A6_26;

!Kovariance reziduí téže položky napříč měřeními
  A1_20-A1_25 Pwith A1_21-A1_26;
  A1_20-A1_24 Pwith A1_22-A1_26;
  A1_20-A1_23 Pwith A1_23-A1_26;
  A1_20-A1_22 Pwith A1_24-A1_26;
  A1_20-A1_21 Pwith A1_25-A1_26;
  A1_20 with A1_26@0;

  A2_20-A2_25 Pwith A2_21-A2_26;
  A2_20-A2_24 Pwith A2_22-A2_26;
  A2_20-A2_23 Pwith A2_23-A2_26;
  A2_20-A2_22 Pwith A2_24-A2_26;
  A2_20-A2_21 Pwith A2_25-A2_26;

```



```

A2_20 with A2_26@0;

A3_20-A3_25 Pwith A3_21-A3_26;
A3_20-A3_24 Pwith A3_22-A3_26;
A3_20-A3_23 Pwith A3_23-A3_26;
A3_20-A3_22 Pwith A3_24-A3_26;
A3_20-A3_21 Pwith A3_25-A3_26;
A3_20 with A3_26@0;

A4_20-A4_25 Pwith A4_21-A4_26;
A4_20-A4_24 Pwith A4_22-A4_26;
A4_20-A4_23 Pwith A4_23-A4_26;
A4_20-A4_22 Pwith A4_24-A4_26;
A4_20-A4_21 Pwith A4_25-A4_26;
A4_20 with A4_26@0;

A5_20-A5_25 Pwith A5_21-A5_26;
A5_20-A5_24 Pwith A5_22-A5_26;
A5_20-A5_23 Pwith A5_23-A5_26;
A5_20-A5_22 Pwith A5_24-A5_26;
A5_20-A5_21 Pwith A5_25-A5_26;
A5_20 with A5_26@0;

A6_20-A6_25 Pwith A6_21-A6_26;
A6_20-A6_24 Pwith A6_22-A6_26;
A6_20-A6_23 Pwith A6_23-A6_26;
A6_20-A6_22 Pwith A6_24-A6_26;
A6_20-A6_21 Pwith A6_25-A6_26;
A6_20 with A6_26@0;

A7_20-A7_25 Pwith A7_21-A7_26;
A7_20-A7_24 Pwith A7_22-A7_26;
A7_20-A7_23 Pwith A7_23-A7_26;
A7_20-A7_22 Pwith A7_24-A7_26;
A7_20-A7_21 Pwith A7_25-A7_26;
A7_20 with A7_26@0;

!Růstová křivka.
  I by A20-A26@1;
  S by A20* (GL1)
    A21 (GL2)
    A22 (GL3)
    A23 (GL4)
    A24 (GL5)
    A25 (GL6)
    A26 (GL7);

  I S;
  I with S;
  [I S];
!Lag-1 kovariance disturbancí faktorů v čase
  A20-A25 pwith A21-A26;
!Konstantní disturbance v čase
  A20-A26 (r);

MODEL CONSTRAINT:
NEW(alpha*.5 lambda*2);
GL1 = 1/(1 + EXP (- (0-lambda)*alpha));
GL2 = 1/(1 + EXP (- (1-lambda)*alpha));
GL3 = 1/(1 + EXP (- (2-lambda)*alpha));
GL4 = 1/(1 + EXP (- (3-lambda)*alpha));
GL5 = 1/(1 + EXP (- (4-lambda)*alpha));
GL6 = 1/(1 + EXP (- (5-lambda)*alpha));
GL7 = 1/(1 + EXP (- (6-lambda)*alpha));

```

```
OUTPUT: SAMPSTAT STDYX RESIDUAL;
        SVALUES;                                !Export parametrů modelu
PLOT: SERIES = A20-A26 (*); TYPE=PLOT3;
```

SO-LCGA model se startovacími hodnotami z CFA modelu

```
TITLE: Unconditional Curve of Factors Model for autonomy
DATA: FILE="AIwide3.dat";
VARIABLE: NAMES = ...
        MISSING = ALL (999);
        USEVAR = A1_20-A7_20 A1_21-A7_21 A1_22-A7_22 A1_23-A7_23
                A1_24-A7_24 A1_25-A7_25 A1_26-A7_26;
        USEOBS = (konec > 20) AND (zacatek < 26) AND (pocaut > 2);
        CLASSES = C(3);
```

```
ANALYSIS: ESTIMATOR=ML;
        COVERAGE=0;
        H1ITERATIONS = 50000;
        H1CONVERGENCE = 0.001;
        PROCESSORS = 12;
        TYPE = MIXTURE;
        STARTS = 80 20;                !Default is 20 4
```

```
MODEL:
%OVERALL%
!Náboje
a20 BY a1_20@1;
a20 BY a2_20*-1.14005 (12);
a20 BY a3_20*0.86204 (13);
a20 BY a4_20*-1.07308 (14);
a20 BY a5_20*0.79026 (15);
a20 BY a6_20*0.84883 (16);
a20 BY a7_20*-1.05027 (17);
a21 BY a1_21@1;
a21 BY a2_21*-1.14005 (12);
a21 BY a3_21*0.86204 (13);
a21 BY a4_21*-1.07308 (14);
a21 BY a5_21*0.79026 (15);
a21 BY a6_21*0.84883 (16);
a21 BY a7_21*-1.05027 (17);
a22 BY a1_22@1;
a22 BY a2_22*-1.14005 (12);
a22 BY a3_22*0.86204 (13);
a22 BY a4_22*-1.07308 (14);
a22 BY a5_22*0.79026 (15);
a22 BY a6_22*0.84883 (16);
a22 BY a7_22*-1.05027 (17);
a23 BY a1_23@1;
a23 BY a2_23*-1.14005 (12);
a23 BY a3_23*0.86204 (13);
a23 BY a4_23*-1.07308 (14);
a23 BY a5_23*0.79026 (15);
a23 BY a6_23*0.84883 (16);
a23 BY a7_23*-1.05027 (17);
a24 BY a1_24@1;
a24 BY a2_24*-1.14005 (12);
a24 BY a3_24*0.86204 (13);
a24 BY a4_24*-1.07308 (14);
a24 BY a5_24*0.79026 (15);
a24 BY a6_24*0.84883 (16);
a24 BY a7_24*-1.05027 (17);
a25 BY a1_25@1;
```

```

a25 BY a2_25*-1.14005 (12);
a25 BY a3_25*0.86204 (13);
a25 BY a4_25*-1.07308 (14);
a25 BY a5_25*0.79026 (15);
a25 BY a6_25*0.84883 (16);
a25 BY a7_25*-1.05027 (17);
a26 BY a1_26@1;
a26 BY a2_26*-1.14005 (12);
a26 BY a3_26*0.86204 (13);
a26 BY a4_26*-1.07308 (14);
a26 BY a5_26*0.79026 (15);
a26 BY a6_26*0.84883 (16);
a26 BY a7_26*-1.05027 (17);
!Průsečíky položek
[ a1_20@0 ];
[ a2_20*9.30379 ] (i2);
[ a3_20*1.11389 ] (i3);
[ a4_20*8.99848 ] (i4);
[ a5_20*0.58006 ] (i5);
[ a6_20*0.15395 ] (i6);
[ a7_20*8.36784 ] (i7);
[ a1_21@0 ];
[ a2_21*9.30379 ] (i2);
[ a3_21*1.11389 ] (i3);
[ a4_21*8.99848 ] (i4);
[ a5_21*0.58006 ] (i5);
[ a6_21*0.15395 ] (i6);
[ a7_21*8.36784 ] (i7);
[ a1_22@0 ];
[ a2_22*9.30379 ] (i2);
[ a3_22*1.11389 ] (i3);
[ a4_22*8.99848 ] (i4);
[ a5_22*0.58006 ] (i5);
[ a6_22*0.15395 ] (i6);
[ a7_22*8.36784 ] (i7);
[ a1_23@0 ];
[ a2_23*9.30379 ] (i2);
[ a3_23*1.11389 ] (i3);
[ a4_23*8.99848 ] (i4);
[ a5_23*0.58006 ] (i5);
[ a6_23*0.15395 ] (i6);
[ a7_23*8.36784 ] (i7);
[ a1_24@0 ];
[ a2_24*9.30379 ] (i2);
[ a3_24*1.11389 ] (i3);
[ a4_24*8.99848 ] (i4);
[ a5_24*0.58006 ] (i5);
[ a6_24*0.15395 ] (i6);
[ a7_24*8.36784 ] (i7);
[ a1_25@0 ];
[ a2_25*9.30379 ] (i2);
[ a3_25*1.11389 ] (i3);
[ a4_25*8.99848 ] (i4);
[ a5_25*0.58006 ] (i5);
[ a6_25*0.15395 ] (i6);
[ a7_25*8.36784 ] (i7);
[ a1_26@0 ];
[ a2_26*9.30379 ] (i2);
[ a3_26*1.11389 ] (i3);
[ a4_26*8.99848 ] (i4);
[ a5_26*0.58006 ] (i5);
[ a6_26*0.15395 ] (i6);
[ a7_26*8.36784 ] (i7);

```

!Průměry faktorů
[A20-A26@0];

!Residua

a1_20*0.80671;
a2_20*1.15111;
a3_20*0.94208;
a4_20*1.22402;
a5_20*1.80309;
a6_20*2.22090;
a7_20*1.03852;
a1_21*0.85112;
a2_21*1.28460;
a3_21*0.81985;
a4_21*1.28312;
a5_21*1.41187;
a6_21*2.07558;
a7_21*1.00969;
a1_22*0.71479;
a2_22*1.35159;
a3_22*0.72314;
a4_22*1.08592;
a5_22*1.46805;
a6_22*1.94191;
a7_22*0.83648;
a1_23*0.64340;
a2_23*1.12888;
a3_23*0.67177;
a4_23*1.11238;
a5_23*1.29073;
a6_23*2.00582;
a7_23*0.75639;
a1_24*0.56995;
a2_24*1.15536;
a3_24*0.58443;
a4_24*0.96096;
a5_24*1.26441;
a6_24*1.72859;
a7_24*0.89943;
a1_25*0.52648;
a2_25*1.27812;
a3_25*0.63191;
a4_25*1.08619;
a5_25*1.09139;
a6_25*1.76630;
a7_25*0.72337;
a1_26*0.47981;
a2_26*1.00907;
a3_26*0.55932;
a4_26*1.04309;
a5_26*1.05765;
a6_26*1.62125;
a7_26*0.89385;

!Kovariance rezidui

a5_20 WITH a6_20*0.22309;
a5_20 WITH a5_21*0.73179;
a5_20 WITH a5_22*0.58656;
a5_20 WITH a5_23*0.55333;
a5_20 WITH a5_24*0.48377;
a5_20 WITH a5_25*0.53278;
a5_20 WITH a5_26@0;
a5_21 WITH a6_21*0.20649;
a5_21 WITH a5_22*0.53293;

a5_21 WITH a5_23*0.46573;
a5_21 WITH a5_24*0.48136;
a5_21 WITH a5_25*0.41524;
a5_21 WITH a5_26*0.47601;
a5_22 WITH a6_22*0.25261;
a5_22 WITH a5_23*0.47644;
a5_22 WITH a5_24*0.54121;
a5_22 WITH a5_25*0.56524;
a5_22 WITH a5_26*0.34889;
a5_23 WITH a6_23*0.24125;
a5_23 WITH a5_24*0.58449;
a5_23 WITH a5_25*0.42545;
a5_23 WITH a5_26*0.38605;
a5_24 WITH a6_24*0.09213;
a5_24 WITH a5_25*0.50493;
a5_24 WITH a5_26*0.44753;
a5_25 WITH a6_25*0.24499;
a5_25 WITH a5_26*0.46350;
a5_26 WITH a6_26*0.12124;
a1_20 WITH a1_21*0.17628;
a1_20 WITH a1_22*0.07608;
a1_20 WITH a1_23*0.09092;
a1_20 WITH a1_24*-0.01024;
a1_20 WITH a1_25*-0.06876;
a1_20 WITH a1_26@0;
a2_20 WITH a2_21*0.31151;
a2_20 WITH a2_22*0.30202;
a2_20 WITH a2_23*0.23147;
a2_20 WITH a2_24*0.32205;
a2_20 WITH a2_25*-0.10188;
a2_20 WITH a2_26@0;
a3_20 WITH a3_21*0.29055;
a3_20 WITH a3_22*0.12173;
a3_20 WITH a3_23*0.19403;
a3_20 WITH a3_24*0.02680;
a3_20 WITH a3_25*0.03038;
a3_20 WITH a3_26@0;
a4_20 WITH a4_21*0.26193;
a4_20 WITH a4_22*0.21118;
a4_20 WITH a4_23*0.36694;
a4_20 WITH a4_24*0.27866;
a4_20 WITH a4_25*0.43614;
a4_20 WITH a4_26@0;
a6_20 WITH a6_21*0.67157;
a6_20 WITH a6_22*0.71827;
a6_20 WITH a6_23*0.80640;
a6_20 WITH a6_24*0.51895;
a6_20 WITH a6_25*0.91889;
a6_20 WITH a6_26@0;
a7_20 WITH a7_21*-0.05358;
a7_20 WITH a7_22*0.09711;
a7_20 WITH a7_23*0.20692;
a7_20 WITH a7_24*0.24615;
a7_20 WITH a7_25*0.11033;
a7_20 WITH a7_26@0;
a1_21 WITH a1_22*0.22360;
a1_21 WITH a1_23*0.11282;
a1_21 WITH a1_24*0.03571;
a1_21 WITH a1_25*0.13044;
a1_21 WITH a1_26*-0.01963;
a2_21 WITH a2_22*0.30134;
a2_21 WITH a2_23*0.26956;
a2_21 WITH a2_24*0.16262;

a2_21 WITH a2_25*0.30905;
a2_21 WITH a2_26*0.27490;
a3_21 WITH a3_22*0.25316;
a3_21 WITH a3_23*0.21764;
a3_21 WITH a3_24*0.18021;
a3_21 WITH a3_25*0.12273;
a3_21 WITH a3_26*0.02524;
a4_21 WITH a4_22*0.29172;
a4_21 WITH a4_23*0.22028;
a4_21 WITH a4_24*0.23970;
a4_21 WITH a4_25*0.26988;
a4_21 WITH a4_26*0.18379;
a6_21 WITH a6_22*0.70917;
a6_21 WITH a6_23*0.64445;
a6_21 WITH a6_24*0.43601;
a6_21 WITH a6_25*0.43323;
a6_21 WITH a6_26*0.28696;
a7_21 WITH a7_22*0.17361;
a7_21 WITH a7_23*0.07043;
a7_21 WITH a7_24*0.11342;
a7_21 WITH a7_25*0.14840;
a7_21 WITH a7_26*-0.05209;
a1_22 WITH a1_23*0.08399;
a1_22 WITH a1_24*0.10669;
a1_22 WITH a1_25*0.00483;
a1_22 WITH a1_26*0.02662;
a2_22 WITH a2_23*0.31224;
a2_22 WITH a2_24*0.23248;
a2_22 WITH a2_25*0.32103;
a2_22 WITH a2_26*0.16685;
a3_22 WITH a3_23*0.21069;
a3_22 WITH a3_24*0.12265;
a3_22 WITH a3_25*0.10241;
a3_22 WITH a3_26*0.10545;
a4_22 WITH a4_23*0.34449;
a4_22 WITH a4_24*0.20285;
a4_22 WITH a4_25*0.23049;
a4_22 WITH a4_26*0.15567;
a6_22 WITH a6_23*0.74703;
a6_22 WITH a6_24*0.51890;
a6_22 WITH a6_25*0.45150;
a6_22 WITH a6_26*0.57870;
a7_22 WITH a7_23*0.14184;
a7_22 WITH a7_24*0.13995;
a7_22 WITH a7_25*0.04311;
a7_22 WITH a7_26*0.07423;
a1_23 WITH a1_24*0.12167;
a1_23 WITH a1_25*0.14931;
a1_23 WITH a1_26*0.09937;
a2_23 WITH a2_24*0.20606;
a2_23 WITH a2_25*0.09579;
a2_23 WITH a2_26*0.12451;
a3_23 WITH a3_24*0.16993;
a3_23 WITH a3_25*0.16188;
a3_23 WITH a3_26*0.11023;
a4_23 WITH a4_24*0.36217;
a4_23 WITH a4_25*0.24360;
a4_23 WITH a4_26*0.32301;
a6_23 WITH a6_24*0.55064;
a6_23 WITH a6_25*0.40311;
a6_23 WITH a6_26*0.58039;
a7_23 WITH a7_24*0.06306;
a7_23 WITH a7_25*0.10499;

```

a7_23 WITH a7_26*-0.06805;
a1_24 WITH a1_25*0.19796;
a1_24 WITH a1_26*0.04048;
a2_24 WITH a2_25*0.32220;
a2_24 WITH a2_26*0.25310;
a3_24 WITH a3_25*0.15933;
a3_24 WITH a3_26*0.15053;
a4_24 WITH a4_25*0.26134;
a4_24 WITH a4_26*0.33566;
a6_24 WITH a6_25*0.25688;
a6_24 WITH a6_26*0.30534;
a7_24 WITH a7_25*0.06801;
a7_24 WITH a7_26*0.01996;
a1_25 WITH a1_26*0.12530;
a2_25 WITH a2_26*0.21250;
a3_25 WITH a3_26*0.25431;
a4_25 WITH a4_26*0.30235;
a6_25 WITH a6_26*0.49352;
a7_25 WITH a7_26*-0.01419;
!Růstová křivka.
I by A20-A26@1;
S by A20* (GL1)
    A21 (GL2)
    A22 (GL3)
    A23 (GL4)
    A24 (GL5)
    A25 (GL6)
    A26 (GL7);
I-S@0;
I with S@0;
[I-S];
!Lag-1 kovariance disturbancí faktorů v čase
A20-A25 pwith A21-A26;
!Konstantní disturbance v čase
A20-A26 (r);

```

```

%c#2%
[I S];

```

```

%c#3%
[I S];

```

```

MODEL CONSTRAINT:
NEW(alpha*.5 lambda*2);
GL1 = 1/(1 + EXP (-(0-lambda)*alpha));
GL2 = 1/(1 + EXP (-(1-lambda)*alpha));
GL3 = 1/(1 + EXP (-(2-lambda)*alpha));
GL4 = 1/(1 + EXP (-(3-lambda)*alpha));
GL5 = 1/(1 + EXP (-(4-lambda)*alpha));
GL6 = 1/(1 + EXP (-(5-lambda)*alpha));
GL7 = 1/(1 + EXP (-(6-lambda)*alpha));

```

```

OUTPUT: SAMPSTAT STDYX;

```

```

PLOT: SERIES = A20-A26 (*); TYPE=PLOT3;

```

SO-GMM model s různými tvary logistické křivky napříč třídami

```

MODEL:
    %OVERALL%
...    !Model měření

    !Růstová křivka.

```

```

I by A20-A26@1;
S by A20* (GL1)
    A21 (GL2)
    A22 (GL3)
    A23 (GL4)
    A24 (GL5)
    A25 (GL6)
    A26 (GL7);
I-S;
I with S;
[I-S];
!Lag-1 kovariance disturbancí faktorů v čase
A20-A25 pwith A21-A26;
!Konstantní disturbance v čase
A20-A26 (r);

%c#2%
[I S];
S by A20* (GL12)
    A21 (GL22)
    A22 (GL32)
    A23 (GL42)
    A24 (GL52)
    A25 (GL62)
    A26 (GL72);

%c#3%
[I S];
S by A20* (GL13)
    A21 (GL23)
    A22 (GL33)
    A23 (GL43)
    A24 (GL53)
    A25 (GL63)
    A26 (GL73);

MODEL CONSTRAINT:
NEW(alpha*.5 lambda*2);
GL1 = 1/(1 + EXP (-(0-lambda)*alpha));
GL2 = 1/(1 + EXP (-(1-lambda)*alpha));
GL3 = 1/(1 + EXP (-(2-lambda)*alpha));
GL4 = 1/(1 + EXP (-(3-lambda)*alpha));
GL5 = 1/(1 + EXP (-(4-lambda)*alpha));
GL6 = 1/(1 + EXP (-(5-lambda)*alpha));
GL7 = 1/(1 + EXP (-(6-lambda)*alpha));
NEW(alpha2*.5 lambda2*2);
GL12 = 1/(1 + EXP (-(0-lambda2)*alpha2));
GL22 = 1/(1 + EXP (-(1-lambda2)*alpha2));
GL32 = 1/(1 + EXP (-(2-lambda2)*alpha2));
GL42 = 1/(1 + EXP (-(3-lambda2)*alpha2));
GL52 = 1/(1 + EXP (-(4-lambda2)*alpha2));
GL62 = 1/(1 + EXP (-(5-lambda2)*alpha2));
GL72 = 1/(1 + EXP (-(6-lambda2)*alpha2));
NEW(alpha3*.5 lambda3*2);
GL13 = 1/(1 + EXP (-(0-lambda3)*alpha3));
GL23 = 1/(1 + EXP (-(1-lambda3)*alpha3));
GL33 = 1/(1 + EXP (-(2-lambda3)*alpha3));
GL43 = 1/(1 + EXP (-(3-lambda3)*alpha3));
GL53 = 1/(1 + EXP (-(4-lambda3)*alpha3));
GL63 = 1/(1 + EXP (-(5-lambda3)*alpha3));
GL73 = 1/(1 + EXP (-(6-lambda3)*alpha3));

```


Příloha 2 – R skripty pro zpracování výstupů z Mplus a tvorbu grafů

SKRIPTY PRO ZPRACOVÁNÍ ANALÝZ RIZIKOVÉHO CHOVÁNÍ

```
##- Init ----
library(MplusAutomation)
library(rhdf5)
library(plyr)
library(readtext)
library(ggplot2)
library(car)
library(mvtnorm)
library(ggthemes)

##- Data ----
setwd("M:/ELSPAC/RCh")
rch <- read.csv("RCH.dat", sep="\t", header = F) # Načtení identického datového souboru, který používá Mplus
names(rch) <- c("id", "n", "pohlavi", "s1_13", "s1_15", "s1_17", "s1_19", "s2ln_13", "s2ln_15", "s2ln_17",
               "s2ln_19", "s3ln_13", "s3ln_15", "s3ln_17", "s3ln_19", "s3dich_13", "s3dich_15", "s3dich_17",
               "s3dich_19", "s4i_13", "s4i_15", "s4i_17", "s4i_19", "s5ln_13", "s5ln_15", "s5ln_17",
               "s5ln_19", "Zimpulsiv", "cap_pos", "cap_neg", "cap_aut", "ZPA_S", "ZPA_R", "ZPA_D", "Zinform",
               "Znormat", "Zdiffus", "filter")
rch[rch==999]<- NA # 999 byl kód pro chybějící hodnoty

# Skupiny pro kreslení lineplotů
rch$S2gr <- recode(round(rch$s2ln_13, digits=0), "c(1,2,3,4) = 1; 0 = 0; 5:hi = 2; else = 3")

##- Individuální růstové křivky ---
S2 <- reshape(rch[,c(1,8:11,39)], idvar = c("id"), varying = c("s2ln_13", "s2ln_15", "s2ln_17", "s2ln_19"),
              direction = "long", times = c(13,15,17,19), sep = "_")
windows(1000,800)
ggplot(na.omit(S2), aes(x= time, y=s2ln, group=id)) +
  facet_wrap(~S2gr)+
  geom_path(size=1, alpha=0.2, show.legend=F, aes(color=id))+
  scale_x_continuous(name="Věk", limits=c(13,19), breaks = c(13,15,17,19))+
  scale_y_continuous(name="Užívání návykových látek S2 (ln*10)", limits=c(0,17))+
  theme_bw(base_size=16, base_family = "serif")

##- S2 LGCM -----

# Příklad zpracování latentního růstového modelu
runModels("M:/ELSPAC/RCh/S2/LGCM", replaceOutfile = "modifiedDate") # Dávkové spuštění modelů

Mplus
S2_LGCM <- readModels("M:/ELSPAC/RCh/S2/LGCM") # Načtení výstupů Mplus
sapply(S2_LGCM, "[", "warnings") # Výpis chybových hlášení

Mplus
S2_LGCM_fity <- rbind.fill(sapply(S2_LGCM, "[", "summaries")) # Shrnutí ukazatelů fitu

# Srovnání modelů
compareModels(S2_LGCM[["rch_s2_lgcm_01.out"]], S2_LGCM[["rch_s2_lgcm_02.out"]],
              show=c("diff", "pdiff", "summaries", "unique"), equalityMargin=c(param=.05, pvalue=.02),
              sort="type", diffTest=TRUE, showNS=FALSE)

compareModels(S2_LGCM[["rch_s2_lgcm_01.out"]], S2_LGCM[["rch_s2_lgcm_03.out"]],
              show=c("diff", "pdiff", "summaries", "unique"), equalityMargin=c(param=.05, pvalue=.02),
              sort="type", diffTest=TRUE, showNS=FALSE)

compareModels(S2_LGCM[["rch_s2_lgcm_03.out"]], S2_LGCM[["rch_s2_lgcm_03b.out"]],
              show=c("diff", "summaries", "unique"), equalityMargin=c(param=.05, pvalue=.02),
              sort="type", diffTest=TRUE, showNS=FALSE)
```

```
compareModels(S2_LGCM[["rch_s2_lgcm_01.out"]], S2_LGCM[["rch_s2_lgcm_04.out"]],
  show=c("diff", "pdiff", "summaries", "unique"), equalityMargin=c(param=.05, pvalue=.02),
  sort="type", diffTest=TRUE, showNS=FALSE)
```

```
#Graf růstové křivky s +-1a2SD interceptu a rozptylu
#Graf neobsahuje reziduální rozptyl - jen rozptyl implikovaný růstovými křivkami
```

```
#Parametry
gpars <- S2_LGCM$`rch_s2_lgcm_04.out`$parameters$unstandardized # Načtení parametrů konkrétního modelu
gmeans <- gpars$est[gpars$paramHeader=="Means"] # Vektor průměrů latentních
koeficientů
gcov <- c(gpars$est[gpars$paramHeader=="Variances"], gpars$est[gpars$paramHeader=="I.WITH"]) #Jejich kovariance
gres <- gpars$est[gpars$paramHeader=="Rezidual.Variances"] # Reziduální rozptyly
gbas <- gpars$est[gpars$paramHeader=="S.|"] # Časové báze
# Funkce pro výpočet rozptylu závislé proměnné implikovaného modelem růstových křivek (věk-13)
varY <- function (t) {varY<-gcov[1]+t^2*gcov[2]+t^2*gcov[3]; varY}

graph_elements<-c("mean", "-1sd", "+1sd", "-2sd", "+2sd") # Jména vykreslovaných křivek

# Grafová data
graphdata <- as.data.frame(graph_elements)
graphdata$sd <- c(0,-1,1,-2,2) # SD pro které se křivky počítají
graphdata$Y_13 <- gmeans[1] + gbas[1]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[1])) # Hodnoty křivek ve 13
graphdata$Y_15 <- gmeans[1] + gbas[2]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[2])) # Hodnoty křivek v 15...
graphdata$Y_17 <- gmeans[1] + gbas[3]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[3]))
graphdata$Y_19 <- gmeans[1] + gbas[4]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[4]))

# Graf samotný
windows(1000,800)
ggplot()+
  scale_x_continuous(name = "Věk", limits = c(13,19), breaks = c(13,15,17,19))+
  scale_y_continuous(name = "Užívání návykových látek S2 (ln*10)", limits = c(-5,25))+
  geom_line(aes(x = c(13,15,17,19), y=t(graphdata[graphdata$sd==0,][,3:6])), color = "black")+
  geom_ribbon(aes(x=c(13,15,17,19), ymin=t(graphdata[graphdata$sd=-1,][,3:6]),
  ymax=t(graphdata[graphdata$sd= 1,][,3:6])), alpha=0.5, fill="blue")+
  geom_ribbon(aes(x=c(13,15,17,19), ymin=t(graphdata[graphdata$sd=-2,][,3:6]),
  ymax=t(graphdata[graphdata$sd=2,][,3:6])), alpha=0.5, fill="blue")+
  ggtitle("S2 Latent base střední růstová křivka +-1SD")+
  theme_bw(base_size=16, base_family = "serif")
```

#- S2 LCG Latent-base fixed across classes -----

Příklad zpracování modelu s latentními třídami růstu (vnitřně homogenními)

```
createModels("M:/ELSPAC/RCh/S2/LCG/S2_LCGtemplate.txt") # Vytvoření sady .inp souborů ze šablony
runModels("M:/ELSPAC/RCh/S2/LCG", replaceOutfile = "modifiedDate")# Jejich dávkové spuštění
S2_LCG <- readModels("M:/ELSPAC/RCh/S2/LCG") # Načtení výstupů
sapply(S2_LCG, "[", "warnings")
```

#Shnutí fitů

```
S2_LCG_fity <- rbind.fill(sapply(S2_LCG, "[", "summaries"))[,12:20]
#nebo lze použít funkci MplusAutomation mixtureSummaryTable(S2_LCG)
```

Shnutí křivkových parametrů více modelů (nested sapply umožní zanořit se do struktury listu)

Sloučení parametrů pod sebe s přidáním identifikace modelu

```
temppar <- sapply(sapply(S2_LCG, "[", "parameters"), "[", "unstandardized") # Extrakce do dočasného listu
S2_LCG_pars <- as.data.frame(temppar[[1]]) # Vytvoření df z prvního prvku
```

```
S2_LCG_pars$model <- substr(names(temppar)[1], 1,
  regexpr("out", names(temppar)[1])-2) # jméno souboru jako ID modelu
```

```
for (i in 2:length(temppar)) { # Připojení parametrů s z ostatních modelů
```

```
  tmp <- as.data.frame(temppar[[i]])
```

```
  tmp$model <- substr(names(temppar)[i], 1, regexpr("out", names(temppar)[i])-2)
```

```
  S2_LCG_pars <- rbind.data.frame(S2_LCG_pars,tmp)
```

```
  rm(tmp)}
```

```
rm(temppar)
```

```
names(S2_LCG_pars)[names(S2_LCG_pars)=="LatentClass"]<-"class" #Přejmenování pro konzistenci napříč tabulkami
```

```

#Jednotlivé modely vedle sebe. Vyřazuji S.E. odhadu - est_se
S2_LCG_pars_w <- reshape(S2_LCG_pars, idvar = c("paramHeader", "param", "class"),
  v.names = c("est","se","pval"),
  timevar = "model", direction="wide", drop="est_se")

#Shrnutí pravděpodobnostních parametrů
tempprob <- sapply(sapply(S2_LCG,"[", "class_counts"), "[", "modelEstimated")
#sloučení pravděpodobnostních parametrů pod sebe s přidáním identifikace modelu
S2_LCG_model_estimated_class_probabilities <- as.data.frame(tempprob[[1]])
S2_LCG_model_estimated_class_probabilities$model <-
  substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_LCG_model_estimated_class_probabilities<-
rbind.data.frame(S2_LCG_model_estimated_class_probabilities,tmp)
  rm(tmp)
}

# Posteriorní pravděpodobnosti členství ve třídě
tempprob <- sapply(sapply(S2_LCG,"[", "class_counts"), "[", "posteriorProb")
S2_LCG_posterior_probabilities <- as.data.frame(tempprob[[1]])
S2_LCG_posterior_probabilities$model <- substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_LCG_posterior_probabilities<-rbind.data.frame(S2_LCG_posterior_probabilities,tmp)
  rm(tmp)
}

# Nejpravděpodobnější třída
tempprob <- sapply(sapply(S2_LCG,"[", "class_counts"), "[", "mostLikely")
S2_LCG_most_likely_class<-as.data.frame(tempprob[[1]])
S2_LCG_most_likely_class$model<-substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_LCG_most_likely_class<-rbind.data.frame(S2_LCG_most_likely_class,tmp)
  rm(tmp)
}
rm(tempprob)

# Grafy pro latent-base křivky
# Parametry pro kreslení
graphdata<-reshape(S2_LCG_pars[(S2_LCG_pars$paramHeader%in%c("Means")&S2_LCG_pars$param%in%c("I","S")
  |S2_LCG_pars$paramHeader%in%c("S.|")),c(2,3,7,8)],
  idvar=c("class", "model"), timevar="param", v.names="est", direction = "wide")
graphdata<-merge(graphdata,S2_LCG_most_likely_class, by=c("class","model")) #připojení četnosti tříd
#Z parametrů je potřeba spočítat body na křivce, protože už nejde o průměr
graphdata$t_13<-graphdata$est.I
graphdata$t_15<-graphdata$est.I + graphdata$est.S*graphdata$est.S2LN_15
graphdata$t_17<-graphdata$est.I + graphdata$est.S*graphdata$est.S2LN_17
graphdata$t_19<-graphdata$est.I + graphdata$est.S*6
#Graf samotný
windows(1000,800)
ggplot(data=reshape(graphdata, idvar=c("model","class"), varying=c("t_13","t_15","t_17","t_19"), direction =
"long", times=c(13,15,17,19), sep="_"))+
  facet_wrap(~model)+
  geom_line(mapping=aes(time,t, col=class, size=proportion))+
  scale_x_continuous(name="Age", limits=c(13,19), breaks = c(13,15,17,19))+
  scale_y_continuous(name="Risk behavior S2 (ln)", limits=c(1,15))+
  scale_linetype(name="s")

# LRT testy - modely se spouští s nejlepším seedem
# Načtení nejlepších seedů, což automat nenačte.

```

```

# Načtení .out jako dlouhých textů
S2_LCG_outy <- readtext::readtext("M:/ELSPAC/RCh/LCG/S2/*.out")
# Nalezení nejlepšího seedu - 6 číslic podle stabilně umístěné fráze "stage start numbers:"
S2_LCG_outy$bestseed <- substr(S2_LCG_outy$text,
                               regexpr("stage start numbers:", S2_LCG_outy$text, perl=TRUE)+44,
                               regexpr("stage start numbers:", S2_LCG_outy$text, perl=TRUE)+49)

# Vyrobení modelů pro BLRT
cat(S2_LCG_outy$bestseed) #pro export seedů do šablony
createModels("M:/ELSPAC/RCh/LCG/S2/S2_LCG_BLRTtemplate.txt") #vytvoření modelů z šablony
runModels("M:/ELSPAC/RCh/LCG", replaceOutfile = "modifiedDate")
# TECH11 a TECH14 readModels nenačte, takže to už musí člověk ručně.

```

#- S2 GMM model -----

Příklad zpracování GMM modelu. Liší se zohledněním variability růstových parametrů v grafech.

```

createModels("M:/ELSPAC/RCh/S2/GMM/S2_GM2template.txt")
runModels("M:/ELSPAC/RCh/S2/GMM", replaceOutfile = "modifiedDate")
S2_GM2<-readModels("M:/ELSPAC/RCh/S2/GMM", filefilter = "GM2")
sapply(S2_GM2, "[", "warnings")
S2_GM2_fity <- rbind.fill(sapply(S2_GM2, "[", "summaries"))[,12:20] #Shrnutí fitů

# Shrnutí křivkových parametrů
temppar<-sapply(sapply(S2_GM2, "[", "parameters"), "[", "unstandardized")
S2_GM2_pars <- as.data.frame(temppar[[1]]) #Sem se to všechno sesype
S2_GM2_pars$model <-
  substr(names(temppar)[1], 1, regexpr("out", names(temppar)[1])-2) #jméno souboru - ID modelu
for (i in 2:length(temppar)) {
  tmp<-as.data.frame(temppar[[i]])
  tmp$model<-substr(names(temppar)[i], 1, regexpr("out", names(temppar)[i])-2)
  S2_GM2_pars<-rbind.data.frame(S2_GM2_pars,tmp)
  rm(tmp)}
rm(temppar)
names(S2_GM2_pars)[names(S2_GM2_pars)=="LatentClass"]<-"class" #Přejmenování kvůli konzistenci
#Jednotlivé modely vedle sebe.
S2_GM2_pars_w<-reshape(S2_GM2_pars, idvar=c("paramHeader", "param", "class"),
                       v.names=c("est", "se", "pval"), timevar="model", direction="wide", drop="est_se")

#Shrnutí pravděpodobnostních parametrů
tempprob<-sapply(sapply(S2_GM2, "[", "class_counts"), "[", "modelEstimated")
S2_GM2_model_estimated_class_probabilities<-as.data.frame(tempprob[[1]])
S2_GM2_model_estimated_class_probabilities$model<-substr(names(tempprob)[1], 1, regexpr("out",
names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_GM2_model_estimated_class_probabilities<-
rbind.data.frame(S2_GM2_model_estimated_class_probabilities,tmp)
  rm(tmp)
}

tempprob<-sapply(sapply(S2_GM2, "[", "class_counts"), "[", "posteriorProb")
#sloučení pravděpodobnostních parametrů pod sebe s přidáním identifikace modelu
S2_GM2_posterior_probabilities<-as.data.frame(tempprob[[1]])
S2_GM2_posterior_probabilities$model<-substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_GM2_posterior_probabilities<-rbind.data.frame(S2_GM2_posterior_probabilities,tmp)
  rm(tmp)
}

tempprob<-sapply(sapply(S2_GM2, "[", "class_counts"), "[", "mostLikely")
#sloučení pravděpodobnostních parametrů pod sebe s přidáním identifikace modelu
S2_GM2_most_likely_class<-as.data.frame(tempprob[[1]])

```

```

S2_GM2_most_likely_class$model<-substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  S2_GM2_most_likely_class<-rbind.data.frame(S2_GM2_most_likely_class,tmp)
  rm(tmp)
}
rm(tempprob)

# Sloučení dvou proměnných identifikujících parametry do jedné proměnné: par
for (i in 1:dim(S2_GM2_pars)[1])
  {S2_GM2_pars$par[i] <- paste(S2_GM2_pars$paramHeader[i], S2_GM2_pars$param[i])}
S2_GM2_pars$par[grep("WITH", S2_GM2_pars$par)]<-"covIS" # přejmenování pro pohodlí

# Grafy pro latent-base GMM křivky
# parametry pro kreslení
graphdata <- reshape(S2_GM2_pars[!(S2_GM2_pars$paramHeader %in%

c("I.", "Intercepts")|S2_GM2_pars$class=="Categorical.Latent.Variables"),c(3,7,8,9)],
  idvar=c("class", "model"), timevar="par", v.names="est", direction = "wide")
graphdata <- merge(graphdata,S2_GM2_most_likely_class, by=c("class", "model")) #připojení četnosti
tříd

# Body na křivce
# Průměrná křivka
graphdata$Y_13<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S. | S2LN_13"
graphdata$Y_15<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S. | S2LN_15"
graphdata$Y_17<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S. | S2LN_17"
graphdata$Y_19<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S. | S2LN_19"

# +1sd křivka (bez reziduálního rozptylu)
graphdata$Yplus_13<-graphdata$Y_13 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_13"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_13"*graphdata$est.covIS")
graphdata$Yplus_15<-graphdata$Y_15 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_15"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_15"*graphdata$est.covIS")
graphdata$Yplus_17<-graphdata$Y_17 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_17"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_17"*graphdata$est.covIS")
graphdata$Yplus_19<-graphdata$Y_19 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_19"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_19"*graphdata$est.covIS")

# -1sd křivka (bez reziduálního rozptylu)
graphdata$Yminus_13<-graphdata$Y_13 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_13"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_13"*graphdata$est.covIS")
graphdata$Yminus_15<-graphdata$Y_15 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_15"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_15"*graphdata$est.covIS")
graphdata$Yminus_17<-graphdata$Y_17 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_17"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_17"*graphdata$est.covIS")
graphdata$Yminus_19<-graphdata$Y_19 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S. | S2LN_19"^2*graphdata$est.Variances S"+
  2*graphdata$est.S. | S2LN_19"*graphdata$est.covIS")

# graf samotný
windows(1000,800)
ggplot(data=reshape(graphdata, idvar = c("model", "class"),
  varying=c("Y_13", "Y_15", "Y_17", "Y_19",
    "Yplus_13", "Yplus_15", "Yplus_17", "Yplus_19",
    "Yminus_13", "Yminus_15", "Yminus_17", "Yminus_19"),
  direction = "long", times=c(13,15,17,19), sep="-")+
  facet_wrap(~model)+
  geom_line(mapping = aes(time, Y, col = class, size = proportion))+
  geom_ribbon(mapping = aes(time, ymin = Yminus, ymax = Yplus, fill = class), alpha = 0.3)+
  scale_x_continuous(name = "Age", limits = c(13,19), breaks = c(13,15,17,19))+
  scale_y_continuous(name = "Risk behavior S2 (ln)", limits = c(0,15))+

```

```
ggtitle("S2 Latent base GMM střední růstové křivky +-1SD")+
theme_bw(base_size = 16, base_family = "serif")
```

SKRIPTY PRO ZPRACOVÁNÍ ANALÝZ AUTONOMIE POMOCÍ SECOND-ORDER GMM NA CURVE OF FACTORS MODELU

```
#- Data ----
setwd("M:/Autonomy/LGM/CFM")

#- Curve of Factor modely -----
runModels("M:/Autonomy/LGM/CFM", replaceOutfile = "modifiedDate", filefilter = "CF")
CFM<-readModels("M:/Autonomy/LGM/CFM", filefilter = "CF1")
sapply(CFM, "[", "warnings")

# Graf růstové křivky s +-1a2SD interceptu a rozptylu
# Graf neobsahuje reziduální rozptyl - jen rozptyl implikovaný růstovými křivkami
# Parametry vybraného modelu
gpars <- CFM$parameters$unstandardized
gmeans <- gpars$est[gpars$paramHeader == "Means"]
gcov <- c(gpars$est[gpars$paramHeader == "Variances"],gpars$est[gpars$paramHeader == "I.WITH"])
gres <- gpars$est[gpars$paramHeader == "Rezidual.Variances"][56] #Jsou konstanční, tak použijí hodnotu z

A26
gbas <- gpars$est[gpars$paramHeader=="S.BY"] #time bases
# Rozptyl závislé v daném věku implikovaný modelem
varY <- function (t) {varY <- gcov[1]+t^2*gcov[2]+t*2*gcov[3]; varY}

# Grafová data
graph_elements <- c("mean", "-1sd", "+1sd", "-2sd", "+2sd")
graphdata <- as.data.frame(graph_elements)
graphdata$sd <- c(0,-1,1,-2,2)
graphdata$Y_20 <- pmin((gmeans[1] + gbas[1]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[1]))),7) # pmin, aby Y<=7
graphdata$Y_21 <- pmin((gmeans[1] + gbas[2]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[2]))),7)
graphdata$Y_22 <- pmin((gmeans[1] + gbas[3]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[3]))),7)
graphdata$Y_23 <- pmin((gmeans[1] + gbas[4]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[4]))),7)
graphdata$Y_24 <- pmin((gmeans[1] + gbas[5]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[5]))),7)
graphdata$Y_25 <- pmin((gmeans[1] + gbas[6]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[6]))),7)
graphdata$Y_26 <- pmin((gmeans[1] + gbas[7]*gmeans[2] + graphdata$sd*sqrt(varY(gbas[7]))),7)

# Graf samotný
windows(1000,800)
ggplot()+
  scale_x_continuous(name ="Věk", limits=c(20,26), breaks = c(20:26))+
  scale_y_continuous(name ="Prožívaná autonomie", limits=c(1,7), breaks = c(1:7))+
  geom_line(aes(x = c(20:26), y = t(graphdata[graphdata$sd == 0,][,3:9])), color="black")+
  geom_ribbon(aes(x = c(20:26), ymin = t(graphdata[graphdata$sd== -1,][,3:9]),
                ymax = t(graphdata[graphdata$sd== 1,][,3:9])), alpha=0.5, fill="blue")+
  geom_ribbon(aes(x = c(20:26), ymin = t(graphdata[graphdata$sd== -2,][,3:9]),
                ymax = t(graphdata[graphdata$sd==2,][,3:9])), alpha=0.5, fill="blue")+
  ggtitle("CFM autonomie, střední růstová křivka +-1SD")+
  theme_bw(base_size=16, base_family = "serif")

#- SO_LCG na CFG s logistickým růstem -----
createModels("M:/Autonomy/LGM/CFM/LC1_template.txt")
runModels("M:/Autonomy/LGM/CFM/", replaceOutfile = "modifiedDate", filefilter = "LCx1" )
SO_LC1 <- readModels("M:/Autonomy/LGM/CFM" , filefilter = "lcx1")
sapply(SO_LC1, "[", "warnings")

# Shrnutí fitů
SO_LC1_fity <- rbind.fill(sapply(SO_LC1, "[", "summaries"))[,12:19]

# Shrnutí křivkových parametrů
```

```

# Sloučení parametrů pod sebe s přidáním identifikace modelu
temppar <- sapply(sapply(SO_LC1,"[", "parameters"), "[", "unstandardized")
SO_LC1_pars <- as.data.frame(temppar[[1]]) # Sem se to všechno sesype
SO_LC1_pars$model <- substr(names(temppar)[1], 1, regexpr("out", names(temppar)[1]) - 2)
for (i in 2:length(temppar)) {
  tmp <- as.data.frame(temppar[[i]])
  tmp$model <- substr(names(temppar)[i], 1, regexpr("out", names(temppar)[i]) - 2)
  SO_LC1_pars <- rbind.data.frame(SO_LC1_pars,tmp)
  rm(tmp)}
rm(temppar)
names(SO_LC1_pars)[names(SO_LC1_pars) == "LatentClass"] <- "class" #Přejmenování kvůli konzistenci napříč
tabulkami

# Jednotlivé modely vedle sebe. Vyřazují est_se
SO_LC1_pars_w <- reshape(SO_LC1_pars, idvar = c("paramHeader", "param", "class"),
  v.names = c("est","se","pval"), timevar = "model", direction = "wide", drop = "est_se")

#Shrnutí pravděpodobnostních parametrů
tempprob<-sapply(sapply(SO_LC1,"[", "class_counts"), "[", "modelEstimated")
SO_LC1_model_estimated_class_probabilities<-as.data.frame(tempprob[[1]])
SO_LC1_model_estimated_class_probabilities$model<-substr(names(tempprob)[1], 1, regexpr("out",
names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  SO_LC1_model_estimated_class_probabilities<-
rbind.data.frame(SO_LC1_model_estimated_class_probabilities,tmp)
  rm(tmp)
}

tempprob<-sapply(sapply(SO_LC1,"[", "class_counts"), "[", "posteriorProb")
SO_LC1_posterior_probabilities<-as.data.frame(tempprob[[1]])
SO_LC1_posterior_probabilities$model<-substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  SO_LC1_posterior_probabilities<-rbind.data.frame(SO_LC1_posterior_probabilities,tmp)
  rm(tmp)
}

tempprob <- sapply(sapply(SO_LC1,"[", "class_counts"), "[", "mostLikely")
#sloučení pravděpodobnostních parametrů pod sebe s přidáním identifikace modelu
SO_LC1_most_likely_class<-as.data.frame(tempprob[[1]])
SO_LC1_most_likely_class$model<-substr(names(tempprob)[1], 1, regexpr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexpr("out", names(tempprob)[i])-2)
  SO_LC1_most_likely_class<-rbind.data.frame(SO_LC1_most_likely_class,tmp)
  rm(tmp)
}
rm(tempprob)

#Grafy pro latent-base křivky
# nejprve vdestilovat parametry pro kreslení
graphdata <- reshape(SO_LC1_pars[(SO_LC1_pars$paramHeader %in% c("Means") & SO_LC1_pars$param %in% c("I","S")
| SO_LC1_pars$paramHeader %in% c("S.BY")), c(2,3,7,8)],
  idvar = c("class", "model"), timevar = "param", v.names = "est", direction = "wide")
graphdata<-merge(graphdata,SO_LC1_most_likely_class, by= c("class","model")) #připojení četnosti tříd

#Body na křivce
graphdata$t_20 <- graphdata$est.I + graphdata$est.S*graphdata$est.A20
graphdata$t_21 <- graphdata$est.I + graphdata$est.S*graphdata$est.A21
graphdata$t_22 <- graphdata$est.I + graphdata$est.S*graphdata$est.A22
graphdata$t_23 <- graphdata$est.I + graphdata$est.S*graphdata$est.A23
graphdata$t_24 <- graphdata$est.I + graphdata$est.S*graphdata$est.A24
graphdata$t_25 <- graphdata$est.I + graphdata$est.S*graphdata$est.A25
graphdata$t_26 <- graphdata$est.I + graphdata$est.S*graphdata$est.A26

```

```

#Graf samotný
windows(1000,800)
ggplot(data = reshape(graphdata, idvar = c("model","class"),
                      varying = c("t_20","t_21","t_22","t_23", "t_24","t_25","t_26"),
                      direction = "long", times = c(20:26), sep = "_")) +
  facet_wrap(~model) +
  geom_line(mapping = aes(time,t, col = class, size = proportion)) +
  scale_x_continuous(name = "Věk", limits = c(20,26), breaks = c(20:26)) +
  scale_y_continuous(name = "Autonomie", limits = c(1,7)) +
  scale_linetype(name = "s")

#- SO-GM2 Různé disturbance napříč třídami-----
createModels("M:/Autonomy/LGM/CFM/GM2_template.txt")
runModels("M:/Autonomy/LGM/CFM", replaceOutfile = "modifiedDate", filefilter = "GM2")
SO_GM2 <- readModels("M:/Autonomy/LGM/CFM", filefilter = "GM2")
sapply(SO_GM2, "[", "warnings")
SO_GM2_fity <- rbind.fill(sapply(SO_GM2, "[", "summaries"))[,12:19] #Shrnutí fitů

# Shrnutí křivkových parametrů
# Sloučení parametrů pod sebe s přidáním identifikace modelu
temppar<-sapply(sapply(SO_GM2, "[", "parameters"), "[", "unstandardized")
SO_GM2_pars<-as.data.frame(temppar[[1]]) #Sem se to všechno sesype
SO_GM2_pars$model<-substr(names(temppar)[1], 1, regexr("out", names(temppar)[1])-2) #jméno souboru jako ID
modelu
for (i in 2:length(temppar)) {
  tmp<-as.data.frame(temppar[[i]])
  tmp$model<-substr(names(temppar)[i], 1, regexr("out", names(temppar)[i])-2)
  SO_GM2_pars<-rbind.data.frame(SO_GM2_pars,tmp)
  rm(tmp)}
rm(temppar)
names(SO_GM2_pars)[names(SO_GM2_pars)=="LatentClass"]<-"class" #Přejmenování kvůli konzistenci napříč
tabulkami
# Jednotlivé modely vedle sebe.
SO_GM2_pars_w <- reshape(SO_GM2_pars, idvar=c("paramHeader", "param", "class"),
                          v.names=c("est","se","pval"), timevar="model", direction="wide", drop="est_se")

# Shrnutí pravděpodobnostních parametrů
tempprob<-sapply(sapply(SO_GM2, "[", "class_counts"), "[", "modelEstimated")
SO_GM2_model_estimated_class_probabilities<-as.data.frame(tempprob[[1]])
SO_GM2_model_estimated_class_probabilities$model<-substr(names(tempprob)[1], 1, regexr("out",
names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexr("out", names(tempprob)[i])-2)
  SO_GM2_model_estimated_class_probabilities<-
rbind.data.frame(SO_GM2_model_estimated_class_probabilities,tmp)
  rm(tmp)
}

tempprob<-sapply(sapply(SO_GM2, "[", "class_counts"), "[", "posteriorProb")
SO_GM2_posterior_probabilities<-as.data.frame(tempprob[[1]])
SO_GM2_posterior_probabilities$model<-substr(names(tempprob)[1], 1, regexr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexr("out", names(tempprob)[i])-2)
  SO_GM2_posterior_probabilities<-rbind.data.frame(SO_GM2_posterior_probabilities,tmp)
  rm(tmp)
}

tempprob<-sapply(sapply(SO_GM2, "[", "class_counts"), "[", "mostLikely")
SO_GM2_most_likely_class<-as.data.frame(tempprob[[1]])
SO_GM2_most_likely_class$model<-substr(names(tempprob)[1], 1, regexr("out", names(tempprob)[1])-2)
for (i in 2:length(tempprob)) {
  tmp<-as.data.frame(tempprob[[i]])
  tmp$model<-substr(names(tempprob)[i], 1, regexr("out", names(tempprob)[i])-2)
}

```



```

SO_GM2_most_likely_class<-rbind.data.frame(SO_GM2_most_likely_class,tmp)
rm(tmp)
}
rm(tempprob)

# Grafy
# Parametry pro kreslení
# Sloučení jména parametrů ze dvou sloupců do proměnné "par"
for (i in 1:dim(SO_GM2_pars)[1])
  {SO_GM2_pars$par[i] <- paste(SO_GM2_pars$paramHeader[i], SO_GM2_pars$param[i])}

graphdata <-
  reshape(SO_GM2_pars[SO_GM2_pars$par %in%
    c("Means S", "Means I", "Variances S", "Variances I", "I.WITH S",
      "S.BY A20", "S.BY A21", "S.BY A22", "S.BY A23", "S.BY A24", "S.BY A25", "S.BY A26"),c(3,7,8,9)],
    idvar=c("class", "model"), timevar="par", v.names="est", direction = "wide")
graphdata <- merge(graphdata,SO_GM2_most_likely_class, by=c("class","model")) #připojení četnosti tříd

# Body na křivce
# Průměrná křivka
graphdata$Y_20<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A20"
graphdata$Y_21<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A21"
graphdata$Y_22<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A22"
graphdata$Y_23<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A23"
graphdata$Y_24<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A24"
graphdata$Y_25<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A25"
graphdata$Y_26<-graphdata$est.Means I" + graphdata$est.Means S"*graphdata$est.S.BY A26"

# Křivka +1sd (bez reziduálního rozptylu)
graphdata$Yplus_20<-graphdata$Y_20 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A20"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A20"*graphdata$est.I.WITH S")
graphdata$Yplus_21<-graphdata$Y_21 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A21"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A21"*graphdata$est.I.WITH S")
graphdata$Yplus_22<-graphdata$Y_22 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A22"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A22"*graphdata$est.I.WITH S")
graphdata$Yplus_23<-graphdata$Y_23 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A23"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A23"*graphdata$est.I.WITH S")
graphdata$Yplus_24<-graphdata$Y_24 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A24"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A24"*graphdata$est.I.WITH S")
graphdata$Yplus_25<-graphdata$Y_25 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A25"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A25"*graphdata$est.I.WITH S")
graphdata$Yplus_26<-graphdata$Y_26 + sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A26"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A26"*graphdata$est.I.WITH S")

# Křivka +-1sd (bez reziduálního rozptylu)
graphdata$Yminus_20<-graphdata$Y_20 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A20"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A20"*graphdata$est.I.WITH S")
graphdata$Yminus_21<-graphdata$Y_21 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A21"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A21"*graphdata$est.I.WITH S")
graphdata$Yminus_22<-graphdata$Y_22 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A22"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A22"*graphdata$est.I.WITH S")
graphdata$Yminus_23<-graphdata$Y_23 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A23"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A23"*graphdata$est.I.WITH S")
graphdata$Yminus_24<-graphdata$Y_24 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A24"^2*graphdata$est.Variances S"+
  2*graphdata$est.S.BY A24"*graphdata$est.I.WITH S")
graphdata$Yminus_25<-graphdata$Y_25 - sqrt(graphdata$est.Variances I"+
  graphdata$est.S.BY A25"^2*graphdata$est.Variances S"+

```

```

2*graphdata$est.S.BY A25"*graphdata$est.I.WITH S")
graphdata$Yminus_26<-graphdata$Y_26 - sqrt(graphdata$est.Variances I"+
graphdata$est.S.BY A26"^2*graphdata$est.Variances S"+
2*graphdata$est.S.BY A26"*graphdata$est.I.WITH S")

# Graf samotný
windows(1000,800)
ggplot(data=reshape(graphdata, idvar=c("model","class"),
varying=c("Y_20","Y_21","Y_22","Y_23","Y_24","Y_25","Y_26" ,
"Yplus_20","Yplus_21","Yplus_22","Yplus_23","Yplus_24","Yplus_25","Yplus_26",
"Yminus_20","Yminus_21","Yminus_22","Yminus_23","Yminus_24","Yminus_25","Yminus_26"),
direction = "long", times = c(20:26), sep = "_")) +
facet_wrap(~model) +
geom_line(mapping=aes(time,Y, col=class, size=proportion)) +
geom_ribbon(mapping=aes(time, ymin=Yminus, ymax=Yplus, fill=class), alpha=0.3) +
scale_x_continuous(name="Věk", limits=c(20,26), breaks = c(20:26)) +
scale_y_continuous(name="Autonomie", limits=c(1,8), breaks = c(2,4,6)) +
ggtitle("Autonomie SO latent-base GMM střední růstové křivky +-1SD") +
theme_bw(base_size=16, base_family = "serif")

```

ŠABLONA PRO GENEROVÁNÍ MPLUS SYNTAXU PRO SADU GMM MODELŮ

Příkaz **createModels** z balíčku **MplusAutomation** generuje sadu **Mplus** syntaxů, které je pak možné dávkově i jednotlivě spouštět. Používá k tomu šablonu, která je vlastně předpříptaveným **Mplus** syntaxem, který je několika jednoduchými příkazy modifikován. Následuje šablona, podle které **createModels** vygeneruje 4 **.inp** soubory se syntaxem **GMM** modelů rizikového chování s jednou až čtyřmi třídami. Tučně jsou zvýrazněny prvky, specifické pro šablonu, které se v syntaxech **Mplus** nevyskytují.

```

[[init]]
iterators = classes;
classes = 1:4;
filename = "S2_GM2_[[classes]].inp";
outputDirectory = "M:/ELSPAC/RCh/S2/GMM";
[[/init]]

TITLE: Unconditional linear GMM for Risk Behavior Scale 2 - substance abuse
DATA: FILE="..\..\RCH.dat";
VARIABLE: NAMES = id n pohlavi s1_13 s1_15 s1_17 s1_19 s2ln_13 s2ln_15 s2ln_17 s2ln_19
s3ln_13 s3ln_15 s3ln_17 s3ln_19 s3dich_13 s3dich_15 s3dich_17 s3dich_19
s4i_13 s4i_15 s4i_17 s4i_19 s5ln_13 s5ln_15 s5ln_17 s5ln_19
Zimpulsiv cap_pos cap_neg cap_aut ZPA_S ZPA_R ZPA_D
Zinform Znormat Zdiffus filter;

MISSING = ALL (999);
USEVAR = s2ln_13 s2ln_15 s2ln_17 s2ln_19;
USEOBSERVATIONS = n > 1;
CLASSES = C([[classes]]);

ANALYSIS:
PROCESSORS = 4;
TYPE = MIXTURE;
ESTIMATOR = MLR;
INFORMATION = OBSERVED;

!Sekce pro hlavní analýzu
STARTS = 800 160;

!Sekce pro LMR test (TECH11) pro BLRT test (TECH14)
!STARTS = 0;

```

```
!OPTSEED=496703;
!K-1STARTS = 80 16;
!LRTBOOTSTRAP = 10;
!LRTSTARTS = 20 4 800 160;
```

```
MODEL:
```

```
%OVERALL%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
s2ln_13-s2ln_19 (e);
```

```
[[classes>1]]
```

```
%c#2%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
```

```
[[classes>2]]
```

```
%c#3%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
```

```
[[classes>3]]
```

```
%c#4%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
```

```
[[classes>4]]
```

```
%c#5%
I S | s2ln_13@0 s2ln_15* s2ln_17* s2ln_19@6;
I-S;
I with S;
```

```
[[/classes>4]]
```

```
[[/classes>3]]
```

```
[[/classes>2]]
```

```
[[/classes>1]]
```

```
OUTPUT: SAMPSTAT STDYX REZIDUAL;
        !TECH11;      !LMR test.
        !TECH14;      !BLRT test.
```

```
PLOT: TYPE=PLOT3;
      SERIES = s2ln_13-s2ln_19(*);
```